INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Algorithms seminar, 1996-1997*

Bruno SALVY, éditeur scientifique

**N ° 3267**
Septembre 1997

THÈME 2


ISSN 0249-6399

# Algorithms seminar, 1996-1997

Bruno SALVY, éditeur scientifique

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n ° 3267 — Septembre 1997 — 140 pages

**Abstract:** These seminar notes represent the proceedings of a seminar devoted to the analysis of algorithms and related topics. The subjects covered include combinatorics, symbolic computation, asymptotic analysis and average-case analysis of algorithms and data structures.

*(Résumé : tsvp)*

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

# Séminaire algorithmes, 1996-1997

**Résumé :** Ces notes de séminaires représentent les actes, en anglais, d'un séminaire consacré à l'analyse d'algorithmes et aux domaines connexes. Les thèmes abordés comprennent : combinatoire, calcul formel, analyse asymptotique et analyse en moyenne d'algorithmes et de structures de données.

# ALGORITHMS SEMINAR 1996–1997

*Bruno Salvy*[1]
*(Editor)*

**Abstract**

These seminar notes represent the proceedings of a seminar devoted to the analysis of algorithms and related topics. The subjects covered include combinatorics, symbolic computation, asymptotic analysis and average-case analysis of algorithms and data structures.

This is the sixth of our series of seminar proceedings. The previous ones have appeared as INRIA Research Reports numbers 1779, 2130, 2381, 2669 and 2992. The content of these proceedings consists of English summaries of the talks, usually written by a reporter from the audience[2].

The primary goal of this seminar is to cover the major methods of the average-case analysis of algorithms and data structures. Neighbouring topics of study are combinatorics, symbolic computation, asymptotic analysis and probabilistic methods.

The study of combinatorial objects—their description, their enumeration according to various parameters—arises naturally in the process of analyzing algorithms that often involve classical combinatorial structures like strings, trees, graphs, and permutations.

Computer algebra plays an increasingly important rôle in this area. It provides a collection of tools that allows one to attack complex models of combinatorics and the analysis of algorithms via *generating functions*; at the same time, it inspires the quest for developing ever more systematic solutions and decision procedures for the analysis of well-characterized classes of problems.

Asymptotic analysis is an essential ingredient in the interpretation of quantitative results supplied by the resolution of combinatorial models. Various asymptotic methods are found to be relevant to the analysis of particular algorithms. These proceedings include singularity analysis, the saddle-point method, Rice's method or Mellin transform techniques.

The thirty-two articles included in this book represent snapshots of current research in these areas. A tentative organization of their contents is given below.

## PART I. COMBINATORIAL MODELS

In addition to its own traditions rooted in mathematics, the study of *combinatorial models* arises naturally in the process of analyzing algorithms that often involve classical combinatorial structures like strings, trees, graphs and permutations.

In [**1**], classical asymptotic methods used in the analysis of algorithms are used to provide information about the shape of unknown generating functions of combinatorial objects. The enumeration of self-avoiding walks in dimension $d$ is a very old open problem of combinatorics. A related simpler problem is solved in [**2**]. A formal language approach to problems related with

[1]This work was supported in part by the Long Term Research Project Alcom-IT (#20244) of the European Union.

[2]The summaries for the past six years are available on the web at the URL `http://www-rocq.inria.fr/algo/seminars/`.

the study of DNA sequences is taken in [3], while [4] uses probabilistic tools to treat one of these problems. The last summary [5] applies the combinatorics of heaps of coins developed in Bordeaux to task resource models.



## PART II. SYMBOLIC COMPUTATION

A generalization of Zeilberger's fast algorithm for definite summation and integration is presented in [6]. A bottleneck in the current implementation of this algorithm is the search for rational solutions of linear systems of equations. An approach to this problem in the differential case is described in [7]. Linear differential operators can be factored by a simple algorithm [8] which relies on differential Galois theory. In [9], differential algebra is used to perform computations on the singular and general solutions of differential algebraic equations. The next summary [10] describes the mathematical tools used in symbolic asymptotics. Some of these tools are applied in [11] to give an algorithm computing the asymptotic behaviour of implicit exp-log functions.



## PART III. ANALYSIS OF ALGORITHMS AND DATA STRUCTURES

The techniques of analytic combinatorics are applied to polynomials over finite fields in [12]. These techniques provide general results on limit distributions of combinatorial parameters which are presented in [13], illustrated by the example of random mappings which has applications in integer factoring algorithms. Similar methods are used in [14] to obtain results about patterns in binary search trees, which are related to the performance of paged binary search trees. The height of binary search trees is discussed by probabilistic methods in [15] while [16] presents a randomized variant of binary search trees which produces trees that are more balanced. The model of binary search trees is also the basis of a nice use of multivariate generating functions [17] for the analysis of multiple quickselect, a variant of the quicksort algorithm which finds simultaneously the $k_1, \ldots, k_r$th elements among $n$ without sorting them. A different model, digital search trees, is studied in [18] for the analysis of the Lempel-Ziv compression algorithms. Tools from dynamical systems are used in [19] for the study of these trees. A combinatorial algorithm is studied in [20] by Brownian motion techniques. Probabilistic tools are also used in [21] and [22].

## PART IV. PROBABILISTIC METHODS

The techniques from analytic combinatorics are complemented by probabilistic methods. For instance, the Wiener-Hopf factorization, presented in **[23]** applies to the study of maxima of random walks. This has applications in computational biology as shown in **[24]**. Markovian models of resource sharing and load transfers are described in **[25,26]**. Relations to number theory are exemplified in **[27]** and **[28]**. Conversely, analytic combinatorics are used in **[29]** to study properties of a distribution. Finally, the Lovász local lemma is used in **[30]** to analyze a graph colouring algorithm.



## PART V. MISCELLANY

Functional analysis and Dirichlet series are used in **[31]** to study the density of reals with constrained continued fractions and a short history of cryptology is the subject of **[32]**.



*Acknowledgements.* The lectures summarized here emanate from a seminar attended by a community of researchers in the analysis of algorithms, from the Algorithms Project at INRIA (the organizers are Philippe Flajolet, François Morain and Bruno Salvy) and the greater Paris area—especially École Polytechnique (Jean-Marc Steyaert), University of Paris Sud at Orsay (Dominique Gouyou-Beauchamps) and LITP (Michèle Soria).

The editor expresses his gratitude to the various persons who actively supported this joint enterprise and offered to write summaries. Thanks are also due to the speakers and to the authors of summaries. Many of them have come from far away to attend one seminar and kindly accepted to write the summary.

We are also greatly indebted to Virginie Collette for making all the organization work smoothly.

The Editor
B. SALVY

# Part 1

# Combinatorial models

# Solvability of Some Combinatorial Problems

*Anthony Guttmann*
Department of Mathematics, University of Melbourne, Australia

December 2, 1996

[summary by Dominique Gouyou-Beauchamps]

**Abstract**

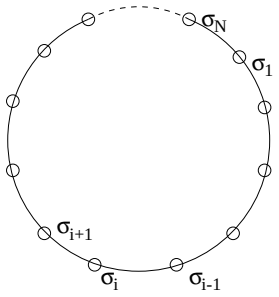
Some of the most famous results in mathematics involve a proof of the intrinsic unsolvability of certain problems—such as the roots of a general quintic. In mathematical physics such results are largely unknown. I will describe a powerful numerical technique that provides compelling evidence for (but is not a proof of) the unsolvability of a wide variety of classical, unsolved problems in Statistical Mechanics and Combinatorics, in terms of the "standard" functions of mathematical physics (including D-finite functions).

## 1. Inversion relation

An inversion relation is a functional relation satisfied by a transfer matrix, and hence satisfied by the partition function (and hence satisfied by correlation functions). It is connected with concepts of integrability and the star-triangle relation (Onsager [7], Baxter [1], Maillard, Stroganoff [8]). An inversion relation exists for many unsolved models e.g., 2d Ising model in a magnetic field, 2d Potts model (non-critical), 3d Ising model (H=0) etc. It gives a strong constraint on the partition function and correlation function.

We first consider the transfer matrix for the Ising model with $d = 1$ and $H \neq 0$ (see [4, 3]). The free energy is

$$-\beta\mathcal{H} = K \sum_{<i,j>} \sigma_i \sigma_j + H \sum_i \sigma_i$$




with $j = i + 1$, $\sigma_i = \pm 1$ and $\sigma_{N+1} = \sigma_1$. The interaction energy between each pair of successive atoms depends on whether their internal coordinates are alike or different: $u(1,1) = u(-1,-1) = -u(1,-1) = -u(-1,1) = -K$. The canonical partition function $Z(K,H)$ of the model is defined as

$$Z(K,H) = \sum_{\sigma} e^{-\beta\mathcal{H}} = \sum_{\sigma_1,\sigma_2,\ldots,\sigma_N=\pm 1} \exp\left(-\sum_k u(\sigma_k,\sigma_{k+1}) - H\sigma_k\right)$$

$$= \sum_{\sigma_1,\sigma_2,\ldots,\sigma_N=\pm 1} \exp\left(\sum_k (K\sigma_k\sigma_{k+1} + H\sigma_k)\right).$$

We express this sum in terms of a transfer matrix $\tilde{\tilde{T}}$ whose elements are

$$(\sigma|\tilde{\tilde{T}}|\sigma') = e^{-u(\sigma,\sigma')+H\delta_{\sigma,\sigma'}} = e^{K\sigma\sigma'+H\delta_{\sigma,\sigma'}}$$

where $\delta$ is the Kronecker delta function. Then the summation over $(\sigma_1, \sigma_2, \ldots, \sigma_N)$ is equivalent to matrix multiplication, and we obtain

$$Z(K,H) = \sum_{\sigma_1} (\sigma_1|\tilde{\tilde{T}}^N|\sigma_1) = \mathrm{Tr}(\tilde{\tilde{T}}^N).$$

The transfer matrix $\tilde{\tilde{T}}$ is

$$\tilde{\tilde{T}} = \begin{pmatrix} e^{K+H} & e^{-K} \\ e^{-K} & e^{K+H} \end{pmatrix}$$

and by inspection

$$\tilde{\tilde{T}}(K,H)\tilde{\tilde{T}}(K+\frac{i\pi}{2}, -H) = 2i\sinh(2K)\tilde{\tilde{1}}$$

and the partition function verifies

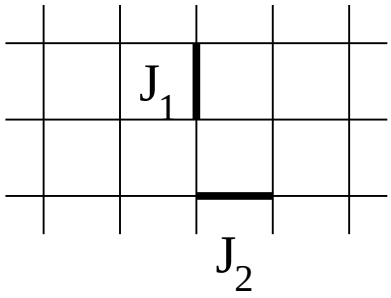
$$Z(K,H)Z(K+\frac{i\pi}{2}, -H) = 2i\sinh(2K).$$

This inversion relation can be confirmed taking the partition function definition:

$$Z(K,H) = e^K \cosh H + (e^{2K}\sinh^2 H + e^{-2K})^{1/2}.$$

For the anisotropic 2d Ising model [7], the free energy is

$$-\beta\mathcal{H} = K_1 \sum_{<i,j>} {}^{(1)}\sigma_i\sigma_j + K_2 \sum_{<i,j>} {}^{(2)}\sigma_i\sigma_j.$$




The diagonal-to-diagonal transfer matrix leads to the following inversion relation for the partition function

$$Z(K_1,K_2)Z(-K_1,K_2+\frac{i\pi}{2}) = 2i\sinh(2K_2).$$

Let $t_1 = \tanh(J_1/kT)$, $t_2 = \tanh(J_2/kT)$. We define the reduced partition function

$$\Lambda(t_1,t_2) = (2\cosh K_1 \cosh K_2)^{-1} Z(K_1,K_2).$$

This satisfies the functional relation

$$\ln\Lambda(t_1,t_2) + \ln\Lambda(1/t_1, -t_2) = \ln(1-t_2^2).$$

Now a series expansions gives:

$$\begin{aligned}\ln\Lambda(t_1,t_2) &= \sum_{m,n} a_{mn} t_1^{2m} t_2^{2n} \\ &= t_1^2t_2^2 + t_1^4t_2^2 + t_1^2t_2^4 + t_1^6t_2^2 + t_1^2t_2^6 + \frac{5}{2}t_1^4t_2^4 + t_1^8t_2^2 + t_1^2t_2^8 + 5t_1^6t_2^4 + 5t_1^4t_2^6 + \cdots\end{aligned}$$

and a partial sum over $m$ yields:

$$\ln\Lambda(t_1,t_2) = \sum_{4n=1}^{\infty} R_n(t_1^2)t_2^{2n}$$

where $R_n(t_1^2) = \sum_m a_{m,n} t_1^{2m}$ is the generating functions for graphs with $2n$ vertical steps. From graphical expansion, we obtain

$$R_1(t_1^2) = \frac{t_1^2}{1-t_1^2}, \qquad R_2(t_1^2) = \frac{t_1^2 - 1/2t_1^4 + 1/2t_1^6}{(1-t_1^2)^3}$$

and from higher order terms we see a pattern [1]

$$R_n(t_1^2) = \frac{P_{2n-1}(t_1^2)}{(1-t_1^2)^{2n-1}}.$$

The fact that the only singularity of the denominator occurs at $t_1^2 = 1$, plus inversion relation, plus obvious symmetry ($\Lambda(t_1, t_2) = \Lambda(t_2, t_1)$) completely determines the polynomials $P_{2n-1}(t^2)$, and hence implicitly yields the Onsager solution [1].

The zero-field susceptibility of the triangular lattice Ising model [5], with coupling constants $K_1$, $K_2$, $K_3$, and $t_i = \tanh(K_i)$, satisfies an inversion relation [6] $\chi(t_1, t_2, t_3) + \chi(-t_1, -t_2, 1/t_3) = 0$. Since the anisotropic square lattice can be obtained by setting one of the anisotropic coupling constants to zero, it follows that the anisotropic square lattice susceptibility satisfies the inversion relation $\chi(t_1, t_2) + \chi(1/t_1, -t_2) = 0$, as well as the symmetry relation $\chi(t_1, t_2) = \chi(t_2, t_1) = 0$. Writing the high temperature expansion

$$\chi(t_1, t_2) = \sum_{m,n=0}^{\infty} c_{mn} t_1^m t_2^n = \sum_{n=0}^{\infty} H_n(t_1) t_2^n,$$

the inversion relation then implies $H_n(t_1) + (-1)^n H_n(1/t_1) = 0$. The first five values of $H_n(x)$ were given in [5], and I. G. Enting and A. J. Guttmann recently reported the calculation of $H_n(x)$ for $n \leq 14$.

$$H_0(t) = \frac{1+t}{1-t}, \qquad H_1(t) = \frac{2(1+t)^2}{(1-t)^2},$$

$$H_2(t) = \frac{2(1+6t+8t^2+6t^3+t^4)}{(1-t)^3(1+t)}, \qquad H_3(t) = \frac{2(1+8t+10t^2+8t^3+t^4)}{(1-t)^4},$$

$$H_4(t) = \frac{2(1+t^{10}+15(t+t^9)+71(t^2+t^8)+192(t^3+t^7)+326(t^4+t^6)+388t^5)}{(1-t^3)(1-t)^4(1+t)^3}, \ldots,$$

$$H_{10}(t) = \frac{2(1+t^{34}+45(t+t^{33})+758(t^2+t^{32})+\cdots+1075878111(t^{16}+t^{18})+1131919146t^{17})}{(1-t^3)^7(1-t)^4(1+t)^9}.$$

In all cases enumerated, the numerator polynomial is unimodal and symmetric with positive coefficients. The numerator and denominator polynomials (with no common factors) are of equal degree. The susceptibility of the Ising model can be expressed as an expansion in terms of $(2k+1)$ particle excitations [9]. Hence the structure of the denominator is clear. The contribution of the terms from the $2k+1$ particle excitations, which first contribute at, say, $O(t^{2m})$, correspond to poles at $t^{2k+1} = 1$ in $H_m(t)$ in the anisotropic expansion above. We can identify the first occurrence of the term $(1-t^{2k+1})$ in the denominator with the first occurrence of a $(2k+1)$ particle excitation. From this observation, we see that structure of $H_n(t)$ is that of a rational function whose poles all lie on the unit circle in the complex $t$ plane, so that poles become dense on the unit circle as $n$ gets large.

Hence $\chi(t_1, t_2)$ as a function of $t_1$ for $t_2$ fixed *(i.)* has a natural boundary; *(ii.)* is not algebraic; and *(iii).* cannot be expressed in terms of the "usual" functions of mathematical physics, such as elliptic integrals, hypergeometric functions etc.

One family of solutions that does suggest itself as a possible candidate is the family of $q$-deformations of standard functions. Several examples are known already in the work done in both Camberra and Melbourne, as well as Bordeaux and various other places.

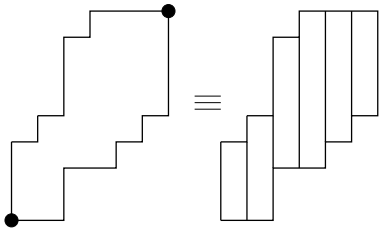
Now this suggests a powerful tool: generalize to the anisotropic model and study the distribution of zeros in the denominator of the analogue of $H_n$ functions. Then we can distinguish between those that appear to be solvable in terms of standard functions—when there is just a finite number of singularities on the unit circle—and those which are not, with an infinite number of such singularities.

## 2. The anisotropic model

We have applied this approach to a number of other unsolved problems in statistical mechanics of two dimensional systems.

**Staircase polygons (or parallelogram polyominoes).** $H_m(x^2)$ is the generating function for staircase polygons with $2m$ vertical bonds.

$$H_1(x^2) = \frac{x^2}{1-x^2}, \qquad H_2(x^2) = \frac{x^2}{(1-x^2)^3},$$
$$H_3(x^2) = \frac{x^2(1+x^2)}{(1-x^2)^5}, \qquad H_4(x^2) = \frac{x^2(1+3x^2+x^4)}{(1-x^2)^7},$$



$$H_5(x^2) = \frac{x^2(1+6x^2+6x^4+x^6)}{(1-x^2)^9}, \ldots, H_m(x^2) = \frac{x^2 S_m(x^2)}{(1-x^2)^{2m-1}}.$$

$S_m(x^2)$ is a symmetric unimodal polynomial of degree $(m-2)$. $H_m(x^2)$ satisfies the inversion relation

$$H_m(x^2) + x^{-2(m-1)} H_m(1/x^2) = 0$$

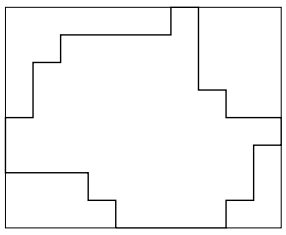
and it follows that the generating function $P(x,y)$ verifies

$$P(x,y) = \frac{x^2 y^2}{1-x^2} + G(x,y) \quad \text{where} \quad G(x,y) = -x^2 G(1/x, y/x) \quad \text{and} \quad P(x,y) = P(y,x).$$

The complete solution is implicitly determined by these relations, initial conditions and the fact that the only singularity of the denominator occurs at $x^2 = 1$.

**Convex polygons.** $H_n(x^2)$ is the generating function for convex polygons with $2n$ vertical bonds.

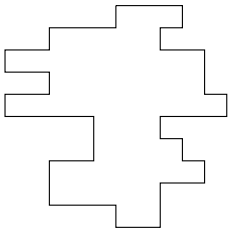
$$H_1(x^2) = \frac{x^2}{1-x^2}, \quad H_2(x^2) = \frac{x^2(1+x^2)^2}{(1-x^2)^3},$$
$$H_3(x^2) = \frac{x^2(1+8x^2+13x^4+2x^6)}{(1-x^2)^5},$$
$$H_n(x^2) = \frac{x^2 T_n(x^2)}{(1-x^2)^{2n-1}}.$$



$T_n$ is a unimodal (with positive coefficients), asymmetric ($n > 2$) polynomial of degree $n$. Absence of symmetry precludes solution by this method.

**Row convex polygons.**

$$P(x,y) \neq P(y,x),$$
$$P(x,y) = \sum_n x^{2n} H_n(y^2) = \sum_m y^{2m} A(x^2).$$



$H_n(y^2) = T_n(y^2)/(1-y^2)^{2n-1}$ where $T_n(y^2)$ is a symmetric, unimodal polynomial of degree $2n-1$ in $y^2$ and $H_n(y^2) = -H_n(1/y^2)$ so $P(x,y) + P(x,1/y) = 0$.

$A_m(x^2) = U_m(x^2)/(1-x^2)^{2m-1}$ where $U_m(x^2)$ is an asymmetric ($m > 2$), unimodal polynomial of degree $2n-1$ in $x^2$. Absence of symmetry precludes solution by this method.

**3 choice polygons.** The number of 3 choice polygons of $2n$ steps grows like

$$\frac{2^{2n-5}}{\sqrt{n\pi}}(15.13161 + \frac{3\sqrt{3}}{\pi}\ln n).$$

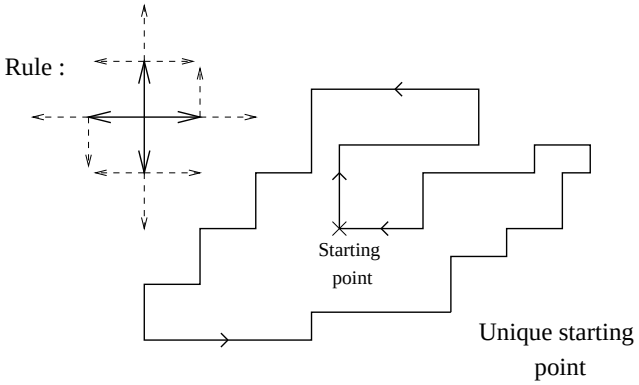
If $H_n(y^2)$ is the generating function for 3 choice polygons with $2n$ horizontal bonds, then the general generating function $P(x,y)$ for 3 choice polygons satisfies

$$P(x,y) = \sum_n x^{2n} H_n(y^2).$$

$$H_2(y^2) = \frac{1}{(1-y^2)^3},$$

$$H_3(y^2) = \frac{3+2y^2}{(1-y^2)^5},$$

$$H_4(y^2) = \frac{6+19y^2+15y^4+4y^6)}{(1-y^2)^7(1+y^2)},$$




$$H_5(y^2) = \frac{10+75y^2+194y^4+237y^6+161y^8+66y^{10}-5y^{12}-12y^{14}+16y^{16}-6y^{20}+2y^{22}}{(1-y^2)^9(1+y^2)^3}.$$

The denominator in general is

$$H_n = (1-y^2)^{2n-1}(1+y^2)^{2n-7} = (1-y^2)^6(1-y^4)^{2n-7} \qquad \text{for } n > 3.$$

This is consistent with solvability and the solution is probably D-finite.

**Square lattice polygons (I. Enting).** We have no inversion relation for this model but $P(x,y) = P(y,x)$. We define $R_n(x^2)$ by the relation

$$P(x,y) = \sum_{n=1}^{\infty} R_n(x^2) y^{2n}.$$

Numerators of the $R_n(x^2)$ are unimodal, positive, but not symmetric. The degree of the numerator is equal to that of the denominator. The denominator is $(1-x^2)^{2n-1}$ for $n \leq 4$, but for $n > 4$ powers of $1-x^4$ enter, and for $n > 6$ we see powers of $1-x^6$ entering, while $n = 8$ marks the first occurrence of powers of $1-x^8$ [3].

**Self avoiding walks on square lattice (A. Conway).** We have no inversion relation for this model. We define $H_n(x)$ by the relation

$$C(x,y) = \sum_{n=0}^{\infty} H_n(x) y^n.$$

$H_n(x)$ (now known up to $n = 12$) is equal to $A_n^m / B_n^m$ where $A_n^m$ and $B_n^m$ are polynomials of degree $m$, $A_n^m$ is unimodal and $B_n^m$ is equal to

$$B_n^m(x) = (1-x)^{\alpha}(1+x)^{\beta}(1+x+x^2)^{\gamma}(1+x^2)^{\delta}.$$

Up to $n = 9$, we have $\alpha = n+1$, $\beta = 2(\lfloor n/2 \rfloor)$ with $n > 1$, $\gamma = n-4$ with $n > 4$ and $\delta = 1, 1, 3$ for $9 \geq n > 6$ [2].

**Honeycomb polygons (brickwork).** $H_n(x^2) = P_n(x^2)/Q_n(x^2)$. The denominator pattern is quite regular. The denominators always have zeros only on the unit circle [3], just at $x^2 = 1$ for $n \leq 3$, with powers of $1 - x^4$ appearing at $n = 4$, powers of $1 - x^6$ entering at $n = 7$ and so on. Numerators are positive coefficients, unimodal, but not symmetric. The numerator and denominator are not of equal degree.

**Anisotropic, hexagonal directed animals (A. J. Guttmann and A. R. Conway).** The generating function is $G(z) = \sum_{n \geq 1} a_n z^n$ where $a_n$ is the number of animals with $n$ sites and one source. For square and triangular lattice Dhar showed

- that the generating function is algebraic;
- that the model is equivalent to hard squares in some sense;
- $a_n \sim \mu^n / \sqrt{n}$ with $\mu = 3$ for the square lattice and $\mu = 4$ for the triangular lattice.

For the hexagonal lattice, Dhar found $\mu = 2.0252 \pm 0.0005$ and no generating function could be found. We extended the series to 99 terms, found $\mu = 2.025131 \pm 0.000005$ but no exact solution. The study of an isotropic model $G(x,y) = \sum_{n \geq 0} H_n(x) y^n$ gives $H_n(x) = N_n(x)/D_n(x)$ where the denominator pattern can be guessed. In this case, we have the symmetry $G(x,y) = G(y,x)$.

## Bibliography


[1] Baxter (R. J.). – Exactly solved models. In Cohen (E. G. D.) (editor), *Fundamental Problems in Statistical Mechanics*. vol. 5. – Amsterdam, 1980. Proceedings of the 1980 Enschede Summer School.

[2] Conway (A. R.) and Guttmann (A. J.). – Square lattice self-avoiding walks and corrections to scaling. *Physical Review Letters*, vol. 77, n° 26, 1996, pp. 5284–5287.

[3] Guttmann (A. J.) and Enting (I. G.). – Inversion relations, the Ising model and self-avoiding polygons. *Nuclear Physics (Proc. Suppl.)*, vol. 47, 1996, pp. 735–738.

[4] Guttmann (A. J.) and Enting (I. G.). – Solvability of some statistical mechanical systems. *Physical Review Letters*, vol. 76, n° 3, 1996, pp. 344–347.

[5] Hansel (D.), Maillard (J. M.), Oitmaa (J.), and Velga (M. J.). – Analytical properties of the anisotropic cubic Ising model. *Journal of Statistical Physics*, vol. 48, n° 1/2, 1987, pp. 69–80.

[6] Jaekel (M. T.) and Maillard (J. M.). – A disorder solution for a cubic Ising model. *Journal of Physics* Series A, vol. 18, 1985, pp. 641–651.

[7] Onsager (L.). – Crystal statistics. I. A two dimensional model with an order-disorder transition. *Physical Review*, vol. 65, 1944, p. 117.

[8] Stroganov (Y. G.). – A new calculation method for partition functions in some lattice models. *Physical Letters*, vol. 74A, n° 1,2, 1979, pp. 116–118.

[9] Wu (T. T.), McCoy (B. M.), Tracy (C. A.), and Barouch (E.). – Spin-spin correlation functions for the two-dimensional Ising model: Exact theory in the scaling region. *Physical Review B*, vol. 13, 1976, pp. 316–374.

# Staircase Polygons, Elliptic Integrals and Heun Functions

*Anthony Guttmann*
Department of Mathematics, University of Melbourne, Australia

December 2, 1996

[summary by Dominique Gouyou-Beauchamps]

**Abstract**

We discuss the perimeter generating function of $d$-dimensional staircase polygons and relate these to the generating function of the square of the $d$-dimensional multinomial coefficients. These are found to satisfy differential equations of order $d-1$. The equations are solved for $d < 5$, and the singularity structure deduced for all values of $d$. The connection with complete elliptic integrals, Heun functions and lattice Green functions is also found. The original article by A. J. Guttmann and T. Prellberg can be found in [4].

## 1. Staircase polygons

Any $d$-dimensional staircase polygon [3, 6] of perimeter $2n$ may be considered as made up of two paths, each of length $n$, with common origin and end point, and with successive steps joining neighbouring points on the lattice $\mathbb{Z}^d$. The two paths are constrained to have no point in common other than the origin and the end point, and successive steps must be in the positive direction in all $d$ coordinates. The number of paths of length $n$ having $k_i$ steps in direction $i$ $(1 \le i \le d)$ is $\binom{k_1+\cdots+k_d}{k_1,\ldots,k_d}$ with $k_1 + \cdots + k_d = n$. Then the number of pairs of such paths is $\binom{n}{k_1,\ldots,k_d}^2$. The generating function $Z_d(x_1,\ldots,x_d)$ for these pairs of paths, including a walk of zero length for later convenience, is

$$Z_d(x_1,\ldots,x_d) = \sum_{k_1,\ldots,k_d=0}^{\infty} \binom{k_1+\cdots+k_d}{k_1,\ldots,k_d}^2 x_1^{2k_1}\cdots x_d^{2k_d}.$$

This generating function produces a chain of staircase polygons, each links of which comprises either a staircase polygon or a double bond. The generating function for double bonds is $\sum_{i=1}^{d} x_i^2$ and we denote the generating function of staircase polygons in $d$ dimensions by $G_d(x_1,\ldots,x_d)$. Let $H(x_1,\ldots,x_d)$ be the generating function for a link

$$H(x_1,\ldots,x_d) = \sum_{i=1}^{d} x_i^2 + 2G_d(x_1,\ldots,x_d).$$

Due to the orientability of walks, each staircase polygon is produced twice in the definition of $H(x_1,\ldots,x_d)$. We set $x_1 = \cdots = x_d = x$. The construction "chain of" corresponds to the functional relation

$$Z(x) = \frac{1}{1 - H(x)}.$$

Hence, by inspection

$$G_d(x) = \frac{1}{2}\left(1 - dx^2 - \frac{1}{Z_d(x)}\right) = \frac{1}{2}\left(1 - dx^2 - \left(\sum_{n=0}^{\infty} S_n^{(d)} x^{2n}\right)^{-1}\right) \tag{1}$$

with

$$S_n^{(d)} = \sum_{k_1+\cdots+k_d=n} \binom{n}{k_1, \ldots, k_d}^2.$$

For $d = 1$ we get $S_n^{(1)} = 1$ and $G_1(x^2) = 0$. For $d = 2$ from the identity $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$ we find

$$G_2(x^2) = \frac{1}{2}\left(1 - 2x^2 - \sqrt{1 - 4x^2}\right).$$

For $d > 2$ we observe the recursion

$$S_n^{(d)} = \sum_{m=0}^{n} \binom{n}{m}^2 S_m^{(d-1)}.$$

By expansion and inspection (aided by computer algebra), we get

$$nS_n^{(2)} = 2(2n-1)S_{n-1}^{(2)}, \qquad n^2 S_n^{(3)} = (10n^2 - 10n + 3)S_{n-1}^{(3)} - 9(n-1)^2 S_{n-2}^{(3)}.$$

These recurrences can be reexpressed as differential equations

$$Z_2'(x) - \frac{2}{1-4x} Z_2(x) = 0,$$

$$Z_3''(x) + \frac{1 - 20x + 27x^2}{x(1-x)(1-9x)} Z_3'(x) - \frac{3(1-3x)}{x(1-x)(1-9x)} Z_3(x) = 0,$$

$$Z_4'''(x) + \frac{3(1 - 30x + 128x^2)}{x(1-4x)(1-16x)} Z_4''(x) + \frac{1 - 68x + 448x^2}{x^2(1-4x)(1-16x)} Z_4'(x) - \frac{4}{x^2(1-4x)} Z_4(x) = 0.$$

These differential equations are all Fuchsian, with regular singular points at the origin, at infinity, and at $x = 1/d^2$, $1/(d-2)^2$, $1/(d-4)^2$, ..., the sequence of singular points terminating at $x = 1$ ($d$ odd) or $x = 1/4$ ($d$ even). Moreover, the solutions that are regular in the neighbourhood of $x = 0$ have singularities with exponents $(d-3)/2$ at the other regular singular points, so that, in particular, the dominant singular behaviour is given by

$$Z_d(x^2) \sim \begin{cases} B_d(1 - d^2x^2)^{(d-3)/2}, & d \text{ even}, \\ B_d(1 - d^2x^2)^{(d-3)/2} \ln(1 - d^2x^2), & d \text{ odd}. \end{cases}$$

where $B_d$ is a constant whose amplitude can be calculated.

## 2. Heun functions and lattice Green Functions

For $d = 3$ the differential equation can be rewritten as Heun's equation [7], a generalization of the ${}_2F_1$ hypergeometric equation to the case of four, rather than three, regular points. We denote the solution as

$$Z_3(x^2) = F\left(\frac{1}{9}, -\frac{1}{3}; 1, 1, 1, 1; x^2\right) = \frac{1}{1 - 9x^2} F\left(\frac{9}{8}, -\frac{3}{4}; 1, 1, 1, 1; \frac{x^2}{x^2 - 1/9}\right).$$

Joyce [5] (and Watson for $P_3(1)$) has shown that this Heun function is related to the simple-cubic lattice Green function

$$P_3(z) = \frac{1}{\pi^3} \iiint_0^\pi \frac{dx_1 dx_2 dx_3}{1 - \frac{z}{3}(\cos x_1 + \cos x_2 + \cos x_3)} \tag{2}$$

through

$$F\left(\frac{9}{8}, -\frac{3}{4}; 1, 1, 1, 1; x_3\right) = (P_3(t))^{\frac{1}{2}} \left(1 - \frac{3}{4}x_1\right)^{-\frac{1}{4}} (1 - x_1)^{\frac{1}{2}} \left(1 - \frac{8}{9}x_3\right)^{-\frac{1}{2}},$$

$$x_3 = \frac{1}{2} + \frac{x_2}{4} - \frac{1}{2}\sqrt{(1 - x_2)(1 - x_2/4)} = \frac{9\omega^2}{9\omega^2 - 1}, \qquad x_2 = \frac{x_1}{x_1 - 1}, \tag{3}$$

$$x_1 = \frac{1}{2} + \frac{x}{6} - \frac{1}{2}\sqrt{(1 - x)(1 - x/9)}, \qquad x = t^2. \tag{4}$$

Further

$$P_3(t) = \left(1 - \frac{3}{4}x_1\right)^{\frac{1}{2}} (1 - x_1)^{-1} \left(\frac{2}{\pi}\right)^2 K(k_+)K(k_-)$$

where

$$k_\pm^2 = \frac{1}{2} \pm \frac{x_2}{4}(4 - x_2)^{\frac{1}{2}} - \frac{1}{4}(2 - x_2)(1 - x_2)^{\frac{1}{2}},$$

and $K(k)$ is the complete elliptic integral of the first kind. Hence we conclude that

$$Z_3^2(\omega^2) = \left(\frac{2}{\pi}\right)^2 (1 - 9\omega^2)^{-1}(1 - \omega^2)^{-1} K(k_+)K(k_-) \tag{5}$$

where the argument of the complete elliptic integral is given implicitly as a function of $\omega$ through equations (2), (3) and (4), and so $G_3(x^2)$ follows immediately from (5) and (1).

Similarly, for $d = 4$, the Heun function can also be transformed [1, 7] to give

$$Z_4(x^2) = \left(F\left(\frac{1}{4}, -\frac{1}{8}; \frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2}; 4x^2\right)\right)^2 = \left(F\left(4, -\frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2}; 16x^2\right)\right)^2$$

$$= \frac{4}{\pi^2} K(k_+)K(k_-). \tag{6}$$

where

$$k_\pm^2 = \frac{1}{2} \pm 8x^2\left(1 - 4x^2\right)^{\frac{1}{2}} - \frac{1}{2}(1 - 8x^2)(1 - 16x_2)^{\frac{1}{2}}.$$

Note that $Z_4(x^2)$ is simply related to the lattice Green function [5] for the face-centered-cubic lattice as

$$P_{f.c.c.}(z) = \frac{3}{3 + z}\left(F\left(4, -\frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2}; \frac{4z}{3 + z}\right)\right)^2$$

and for the diamond lattice as

$$P_{diam}(z) = \left(F\left(4, -\frac{1}{2}; \frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2}; z^2\right)\right)^2.$$

Finally, $G_4(x^2)$ follows from (1) and (6).

For $d > 4$ the theory of generalized hypergeometric functions with five or more regular singular point is not known to us, though the full singularity structure of the differential equations is given for $d = 5$ and $d = 6$, and is readily constructible for other values of $d$.

## 3. Bessel functions

Bessel functions [8] are the solutions of the differential equations

$$z^2 f''(z) + z f'(z) + (z^2 - \nu^2) f(z) = 0. \tag{7}$$

Let $J_\nu(z)$ be a solution of (7) which is analytic near the origin. Then we have

$$J_\nu(z) = \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{1}{2}z\right)^{\nu+2m}}{m!\Gamma(\nu+m+1)}$$

when $2\nu$ is not an integer. $J_{-\nu}(z)$ is also such a solution. The first of the two series defines a function called a Bessel function of order $\nu$ and argument $z$, of the first kind. When $\nu$ is an integer, the function

$$Y_n(z) = \lim_{\nu \to n} \frac{J_\nu(z) - (-1)^n J_{-\nu}(z)}{\nu - n}$$

is called a Bessel function of the second kind of order n. Note that

$$Y_n(z) = \left[\frac{\partial J_\nu(z)}{\partial \nu} - (-1)^n \frac{\partial J_{-\nu}(z)}{\partial \nu}\right]_{\nu=n}.$$

It has seemed desirable to Nielsen to regard the pair of the functions $J_\nu(z) \pm iY_\nu(z)$ as standard solutions of Bessel's equation (7). In honour of Hankel, Nielsen denotes them by the symbol $H$

$$H_\nu^{(1)} = J_\nu(z) + iY_\nu(z), \qquad H_\nu^{(2)} = J_\nu(z) - iY_\nu(z).$$

Physicists consider the equation

$$z^2 f''(z) + z f'(z) - (z^2 + \nu^2) f(z) = 0. \tag{8}$$

The solutions of this equation are called modified Bessel functions. It is usually desirable to present the solution of (8) in a real form, and the fundamental systems $J_\nu(iz)$ and $J_{-\nu}(iz)$ or $J_\nu(iz)$ and $Y_\nu(iz)$ are unsuited for this purpose. The function $e^{-\frac{1}{2}\nu i\pi} J_\nu(iz)$ is real in $z$ and is a solution of (8). It is customary to denote the modified Bessel function of the first kind by $I_\nu(z)$ so that

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{\left(\frac{1}{2}z\right)^{\nu+2m}}{m!\Gamma(\nu+m+1)}.$$

The modified Bessel function of the second kind is

$$K_n(z) = \lim_{\nu \to n} \frac{(-1)^n}{2} \left(\frac{I_\nu(z) - I_{-\nu}(z)}{\nu - n}\right).$$

Note that

$$K_\nu(z) = \frac{\pi}{2} \left(\frac{I_{-\nu}(z) - I_\nu(z)}{\sin(\nu\pi)}\right).$$

The following formulæ are valid:

$$(m!)^2 = \int_0^\infty t(t/2)^{2m} K_0(t)\, dt, \qquad e^{-s} = \frac{2s}{\pi} \int_0^1 \frac{K_0(s/u)}{u^2\sqrt{1-u^2}}\, du,$$

$$I_0(z) = \sum_{n=0}^{\infty} \frac{z/2)^{2n}}{(n!)^2} = \frac{1}{\pi} \int_0^\infty e^{z\cos\theta}\, d\theta, \qquad \frac{1}{z} = \int_0^\infty e^{-sz}\, ds. \tag{9}$$

Philippe Flajolet remarks that a natural tool to attack the calculation of $G_d(z)$ are Bessel generating functions [2]. The Bessel generating function of a sequence $s_n$ of numbers is defined as the sum

$$\sum_{n=0}^{\infty} \frac{s_n z^n}{(n!)^2}.$$

For instance, the Bessel generating function of the constant sequence $s_n = 1$ is given by the modified Bessel function

$$j(z) := I_0\left(2\sqrt{z}\right) = \sum_{n=0}^{\infty} \frac{z^n}{(n!)^2}.$$

The Bessel generating function of the number of pairs of paths with same origin and end and positive steps in a $d$-dimensional lattice is given by the product

$$\prod_{i=1}^{d} j(x_i),$$

where $x_i$ marks the steps in dimension $i$. It follows that the Bessel generating function of the numbers $S_n^{(d)}$ with respect to the length $n$ of the paths is $j(x)^d$. In dimension 2, the generating function $Z_2(x)$ is algebraic. In higher dimension, $Z_d(x)$ belongs to the larger class of holonomic functions. This class is closed under sum, product, Borel and inverse Borel transforms. The Borel transform is related to the inverse Laplace transform and is defined by

$$\text{Borel}\left(\sum_{n=0}^{\infty} s_n x^n\right) = \sum_{n=0}^{\infty} \frac{s_n x^n}{n!}.$$

We proceed to get $Z_d(x)$ by the formula

$$Z_d(x) = \left(\text{Borel}^{(-2)}\right)\left(j(x)^d\right).$$

## 4. The 4D simple-cubic lattice Green function

This section is due to the help of L. Glasser. The 4D simple-cubic lattice Green function is

$$\begin{aligned} P_4(z) &= \frac{1}{\pi^4} \iiiint_0^{\pi} \frac{dx_1 dx_2 dx_3 dx_4}{1 - \frac{z}{4}(\cos x_1 + \cos x_2 + \cos x_3 + \cos x_4)} \\ &= \frac{1}{\pi^4} \int_0^{\infty} e^{-s} \left(\int_0^{\pi} e^{zs\cos k}\, dk\right)^4 ds = \int_0^{\infty} e^{-s} I_0^4(sz)\, ds. \end{aligned}$$

From (9)

$$\begin{aligned} P_4(z) &= \sum_{n_1,n_2,n_3,n_4=0}^{\infty} \frac{(z/2)^{2(n_1+n_2+n_3+n_4)}}{(n_1!n_2!n_3!n_4)^2} \int_0^{\infty} e^{-s} s^{2(n_1+n_2+n_3+n_4)}\, ds \\ &= \sum \{n_i\} \frac{2\,(\sum n_i)!}{(n_1!n_2!n_3!n_4)^2} (z/2)^{2\sum n_i} = \sum_{n=0}^{\infty} A_n (z/2)^{2n}, \end{aligned}$$

where $A_n = \sum_{n_1+n_2+n_3+n_4=n} \frac{(2n)!}{(n_1!n_2!n_3!n_4)^2}$. But if we remark that

$$\sum_{n_1+n_2+n_3+n_4=n} (n_1!n_2!n_3!n_4)^{-2} = 2^{2n} \sum_{n_1+n_2=n} \frac{(1/2)_{n_1}(1/2)_{n_2}}{(n_1!n_2!)^3} = 2^{2n} S_n$$

where $(1/2)_{n_1} = \Gamma(n_1 + 1/2)/\Gamma(1/2)$, then we have

$$P_4(z) = \int_0^\infty e^{-s} I_0^4(sz)\, ds = \sum_{n=0}^\infty z^{2n}(2n)! \sum_{n_1+n_2=n} \frac{(1/2)_{n_1}(1/2)_{n_2}}{(n_1!n_2!)^3}.$$

Using the following identities

$$(1/2)_{n-k} = \Gamma(n-k+1/2)/\Gamma(1/2) = \frac{(-1)^{n-k}\pi}{\Gamma(1/2)\Gamma(1/2-n)(1/2-n)_k} \text{ and } (n-k)! = \frac{(-1)^k n!}{(-n)_k}$$

we obtain

$$S_n = \frac{(-1)^n \pi^{1/2}}{\Gamma(1/2-n)(n!)^3} \sum \frac{(1/2)_k (-n)_k (-n)_k (-n)_k}{(1)_k (1)_k (1/2-n)_k k!} = \frac{(-1)^n \pi^{1/2}}{\Gamma(1/2-n)(n!)^3}\, {}_4F_3\left[\begin{matrix} 1/2, -n, -n, -n; \\ 1, 1, 1/2-n; \end{matrix}\ 1\right]$$

and

$$A_n = \frac{4^n (2n)! (1/2)_n}{(n!)^3}\, {}_4F_3\left[\begin{matrix} 1/2, -n, -n, -n; \\ 1, 1, 1/2-n; \end{matrix}\ 1\right].$$

We can also solve the fourth order differential equation with 4 regular singular points (0, 1/16, 1/4, $\infty$). An indirect method leads to an integral representation whereas a direct method leads to Kampé de Fériet functions.

For $d > 4$, we have

$$P_d(z) = \frac{1}{\pi^d} \int \cdots \int_0^\pi \frac{dx_1 \cdots dx_d}{1 - \frac{z}{d}(\cos x_1 + \cdots + \cos x_d)} = \int_0^\infty e^{-s} I_0(sz/d)\, ds = \frac{2}{\pi} \int_0^1 \frac{Z_d(u^2 z^2/d^2)}{\sqrt{1-u^2}}\, du$$

$$\text{with } Z_d(x) = \int_0^\infty t K_0(t) I_0^d(xt)\, dt.$$

## Bibliography


[1] Appell (M.). – *Comptes-Rendus de l'Académie des Sciences*, vol. 91, 1880, pp. 211–214.

[2] Chyzak (Frédéric). – Staircase polygons, a simplified model for self-avoiding walks. – 1997. Available at the URL `http://www-rocq.inria.fr/algo/libraries/autocomb/autocomb.html`.

[3] Flajolet (Philippe). – *Pólya Festoons.* – Research report, INRIA, July 1991. 7 pages.

[4] Guttmann (A. J.) and Prellberg (T.). – Staircase polygons, elliptic integrals, Heun functions, and lattice Green functions. *Physical Review E*, vol. 47, n° 4, April 1993, pp. 2233–2236.

[5] Joyce (G. S.). – On the simple cubic lattice Green function. *Philosophical Transactions of the Royal Society of London. Series A.*, vol. 273, n° 1236, 1973, pp. 583–610.

[6] Pólya (G.). – On the number of certain lattice polygons. *Journal of Combinatorial Theory,* Series A, vol. 6, 1969, pp. 102–105.

[7] Snow (Chester). – *Hypergeometric and Legendre functions with applications to integral equations of potential theory.* – U.S. Government Printing Office, Washington, D. C., 1952, *National Bureau of Standards Applied Mathematics Series*, vol. 19.

[8] Watson (G. N.). – *A Treatise on the Theory of Bessel Functions*, Chapter 2. – Cambridge University Press, Cambridge, 1944, 2nd edition.

# Generating Functions in Computational Biology

*Mireille Régnier*
Algorithms Project — Inria

March 3, 97

[summary by Mireille Régnier]

**Abstract**

We present a few enumeration problems that arose in computational biology. We point out on several examples how symbolic enumeration methods allow for simplifying and extending previous results. We also present some asymptotics.

## 1. Secondary Structures

As a first example, we enumerate here combinatorial structures associated to RNA sequences, e.g., secondary structures, hairpins, cloverleaves, ... ). This study has been started in [11, 13, 8, 4, 12] with inductions on the size of the RNA sequences. We show here that symbolic enumeration methods allow to find directly equations on generating functions and extend previous results. More precisely, these inductions may deeply depend on the values of the parameters involved such as the minimal size $h$ of the helices, the minimal size $b$ of the loops, ... Due to the number of initial conditions, the recurrence relations may become very intricate. We avoid this unnecessary step.

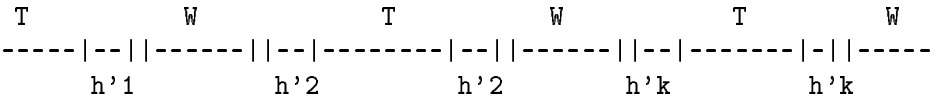
A secondary structure is a set of *helices*, i.e., paired subsequences of the same size. Two paired subsequences are separated by an embedded secondary structure or by a non-paired subsequence, called a *loop*. Secondary structures satisfy three additional conditions. First, the helices must have a minimal size, $h$. Second, the loops must have a minimal size, $b$. Third, two helices cannot overlap.

We now count the number of different secondary structures that may be built on the RNA sequences of a given size:

THEOREM 1. *Let $\mathcal{S}_n^{[h,b]}$ be the set of RNA secondary structures of size $n$, where the minimal helix size and loop size are $h$ and $b$. The generating function $S^{[h,b]}(z)$satisfies the* second degree *equation:*

$$(S^{[h,b]})^2(z)z^{2h} + S^{[h,b]}(z)[(z-1)(1-z^2+z^{2h}) - z^{2h}\frac{1-z^b}{1-z}] + 1 - z^2 + z^{2h} = 0. \quad (1)$$

We consider the external helices, and get $S^{[h,b]}$ from the following decomposition:




where $\mathcal{W}$ is the set of of non-paired subsequences, and $\mathcal{T}^{[h,b]}$ the set of secondary structures that do not start or end with a pairing. Otherwise, such a pairing would be part of the external helix. In

this scheme, we assume there exist $k$ external helices of sizes $h'_1, \ldots, h'_k$. Two external helices are separated by non-paired positions, which is coded by language $\mathcal{W}$. We get the set decomposition:

$$\mathcal{S}^{[h,b]} = \mathcal{W} \times \cup_{k\geq 0}[\mathcal{A}^{[h]} \times \mathcal{T}^{[h,b]} \times \mathcal{W}]^k \tag{2}$$

where $\mathcal{A}^{[h]}$ is the set of couples of subsequences of the same size $h' \geq h$. One can prove a simple relation between $\mathcal{T}^{[h,b]}$ and $\mathcal{S}^{[h,b]}$. Namely:

$$\mathcal{S}^{[h,b]} = \mathcal{A}^{[h]} \times \mathcal{T}^{[h,b]} + \mathcal{T}^{[h,b]} + \mathcal{Y}^{[b]} \tag{3}$$

where $\mathcal{Y}^{[b]}$ counts sequences of length smaller than $b$ (no structure is possible). I.e. $Y^{[b]}(z) = \sum_{i=0}^{b-1} z^i$. We plug (3) into (2) and, applying translation rules, we get (1) after simplification.

REMARK. For $h = 1$ and $b = 1$, we get the equation in [10].

The next theorem directly follows from Darboux's theorem and equation (1).

THEOREM 2. *When* $n \longrightarrow \infty$,

$$S_n^{[h,b]} \sim \frac{1}{4\sqrt{\pi n^3}}[-\rho\Delta'_{h,b}(\rho)]^{\frac{1}{2}} \cdot \rho_{h,b}^{-(n+2h)} \tag{4}$$

*where* $\rho$ *is the smallest positive root of the discriminant* $\Delta_{h,b}(z)$ *of (1).*

REMARK. The location of the roots is discussed in [9]. Notably, it is proven that $1/\rho_{h,b}$ decreases when $h$ and $b$ increase and that, for any $h$ and $b$, $\triangle_{h,b}(z)$ has one root $\rho_{h,b}$ in $]0,1[$ and that $\Delta_{h,b}(z)$ has one root in $]0,1/2[$. It follows that $S_n^{[h,b]}$ grows exponentially and that $S_n^{[1,b]} \geq 2^n$. Numerical values of $\rho$ have been computed using Maple.

One also derives functional equations satisfied by the generating functions of specific structures: the so-called *hairpins* and *cloverleaves*. Asymptotics follow similarly.

## 2. Counting alignments

### 2.1. Language decomposition.

DEFINITION 1. An alignment of $k$ sequences of size $(n_i)_{i=1,k}$ is a sequence of $k$-tuples

$$(\alpha_1^{[j]}, \ldots, \alpha_i^{[j]}, \ldots, \alpha_k^{[j]})$$

such that:

- each subsequence $(\alpha_i^{[j]})$ is increasing from 0 to $n_i$;
- two successive $k$-tuples are not equal.

An aligned position is a an integer $j$ such that

$$\alpha_i^{[j]} = \alpha_i^{[j-1]} + 1, \qquad i = 1, \ldots, k.$$

A $b$-block is a subsequence of aligned positions of length at least $b$.

The number of $k$-alignments is counted in [2]. A first generalization is proposed by [3], for $k = 2$. Blocks of size below some threshold $b$ are considered as non-significant, and one eliminates all alignments containing such small blocks. Nevertheless, the authors still count separately non-aligned letters. E.g. $\begin{smallmatrix} C & \\ & G \end{smallmatrix} \neq \begin{smallmatrix} & C \\ G & \end{smallmatrix}$. We extend this counting for any $k \geq 2$.

A second generalization is proposed in [11] for $k = 2$ and $b = 1$. One identifies $\begin{smallmatrix} C & \\ & G \end{smallmatrix}$ and $\begin{smallmatrix} & C \\ G & \end{smallmatrix}$. We extend it for matching blocks of size $b$ greater than 1. The motivation is the following. In [3], one considers only matching blocks of size at least $b$ as significant. It follows that the difference

between $\begin{smallmatrix} C & - \\ - & G \end{smallmatrix}$ and $\begin{smallmatrix} - & C \\ G & - \end{smallmatrix}$ should be considered as non significant. Hence, in this section, we identify all alignments that differ by a set of positions without a $b$ matching block. When $b = 1$ and $k = 2$, this is the generalization described in [11].

THEOREM 3. *Let $f(n_1, \ldots, n_k)$ be the number of alignments between $k$ sequences of lengths $n_1, \ldots, n_k$. The associated multivariate generating function is, in the ordered case:*

$$\phi(z_1, \ldots, z_k) = \frac{1 - t + t^b}{1 - s(1 - t + t^b) - t} \tag{5}$$

*where $t = \prod_{i=1}^{k} z_i$ and $s = \sum_{i=1}^{k} z_i$. The bivariate generating function is, in the ordered case:*

$$\phi(z_1, \ldots, z_k) = (1 - t + t^b)\frac{1}{1 - (p-1)(1 - t + t^b) - t} \tag{6}$$

*where $p = \prod_{i=1}^{k}(1 - z_i)$.*

When $b = 1$ and $k = 2$, this simplifies into $\phi(z_1, z_2) = 1/[1 - (z_1 + z_2)]$. It follows that $g_1(n, m) = \binom{n+m}{n}$ which is the result proved in [3] by a combinatorial approach.

The proof relies on the derivation of a coding language for sequences $f(n_1, \ldots, n_k)$. For example, one proves, for $b = 1$, that $\mathcal{L}_1 = [\prod_{i=1}^{k}(1 + Z_i) - 1]^*$ [2]. Let us build an alignment from left to right. In each position, one may choose, for each sequence, either to align one character or to introduce a gap. This contributes either by $Z_i$ or by 1, and we get, for independent choices, $\prod_{i=1}^{k}(1 + Z_i)$. The only choice to be excluded is the choice of a gap in all sequences, which is associated to $1^k = 1$. Hence, we get Equation (5). It is worth noticing this is a bivariate function of $s$ and $t$.

**2.2. Asymptotics.** A possible, and usual, assumption is that aligned sequences have the same size. Hence, one is interested in

$$f(n) = [z_1^n \cdots z_k^n]\phi(z_1, \ldots, z_k).$$

The generating function $\sum_n f(n)z^n$ is called the *diagonal* of the generating function $\phi(z_1, \ldots, z_k)$. One can find general results on diagonals in [1, 5]. We provide here a simplified scheme of the approach in the particular case of the alignment of two sequences. We prove:

THEOREM 4. *Let us define:*

$$\Delta_1(t) = (1 - t)^2 - 4t(1 - t + t^b)$$
$$\Delta_2(t) = (1 - t^2 + t^{b+1}) - 4t(1 - t + t^b).$$

*We have $f(n) \sim \alpha n^{-1/2}\rho^{-n}$ where $\rho$ is the (positive) root of smallest modulus of $\Delta_1(t)$ (respectively $\Delta_2(t)$) in the non-ordered (respectively ordered) case.*

PROOF. When $k = 2$, $s$ and $p$ depend only on two variables, $t$ and $z_1$. Namely, one has: $s = (z_1 + t/z_1)$ and $p = 1 + t + z_1 + t/z_1$. In both cases, one has:

$$F(t) = \sum_n f(n)t^n = [z_1^0]f(z_1, t/z_1).$$

Applying the Cauchy formula, we get:

$$F(t) = \frac{1}{2i\pi}\int \frac{f(z_1, t/z_1)}{z_1}\, dz_1 = \frac{1}{2i\pi}\int \frac{\psi(t)}{P(t, z_1)}\, dz_1$$

where $P(t, x)$ is a polynomial in $t$ and $x$ of degree 2 with respect to the second variable $x$. One proves that the discriminant of $P(t, x)$ with respect to $x$ is $\Delta_1(t)$ (respectively $\Delta_2(t)$) in the non-ordered (respectively ordered) case. We first compute the integral. Then, applying Darboux's

theorem, we get that $f(n)/\rho^n$ has a polynomial growth, where $\rho$ is the smallest positive root of $\delta_1(t)$ (respectively $\Delta_2(t)$). Darboux's theorem also provides the dominating term of $f(n)/\rho^n$. We refer the reader to [11] where this term is explicitly given for the non-ordered case. □

## 3. Miscellaneous problems

Many other parameters of interest to biologists can be studied through a generating function approach. One can cite the longest runs [11] or filtration methods such as the statistical distance [6]. This talk presented new results for the statistical distance in a non-uniform probability model. Finally, statistics for the number of occurrences of a given set of words is also of great interest. We presented a generating function approach [7]. We proved the limiting distribution is asymptotically normal and provided formulæ to compute the moments or the probability of occurrences in the finite range, for Bernoulli and Markov models.

## Bibliography


[1] Furstenberg (Harry). – Algebraic functions over finite fields. *Journal of Algebra*, vol. 7, 1967, pp. 271–277.

[2] Griggs (J. R.), Hanlon (P.), Odlyzko (A. M.), and Waterman (M. S.). – On the number of alignments of $k$ sequences. *Graphs and Combinatorics*, vol. 6, n° 2, 1990, pp. 133–146.

[3] Griggs (Jerrold R.), Hanlon (Philip J.), and Waterman (Michael S.). – Sequence alignments with matched sections. *SIAM Journal on Algebraic Discrete Methods*, vol. 7, n° 4, 1986, pp. 604–608.

[4] Howell (J. A.), Smith (T. F.), and Waterman (M. S.). – Computation of generating functions for biological molecules. *SIAM Journal on Applied Mathematics*, vol. 39, n° 1, 1980, pp. 119–133.

[5] Litow (B.) and Dumas (Ph.). – Additive cellular automata and algebraic series. *Theoretical Computer Science*, vol. 119, n° 2, October 1993, pp. 345–354.

[6] Pevzner (P. A.). – Statistical distance between texts and filtration methods in sequence comparison. *CABIOS*, vol. 8, n° 2, 1992, pp. 121–127.

[7] Régnier (M.) and Szpankowski (W.). – On the approximate pattern occurences in a text, 1997. In Proceeding SEQUENCE'97, Positano.

[8] Stein (P. R.) and Waterman (M. S.). – On some new sequences generalizing the Catalan and Motzkin numbers. *Discrete Mathematics*, vol. 26, n° 3, 1978, pp. 261–272.

[9] Tahi (F.). – *Méthodes formelles d'analyse des séquences de nucléotides.* – Thèse de 3e cycle, Université de Paris XI, Orsay, 1997. 155 pages.

[10] Viennot (X. G.) and Vauchaussade de Chaumont (M.). – Enumeration of RNA's secondary structures by complexity. In Capasso (V.), Grosso (E.), and Paven-Fontana (S. L.) (editors), *Mathematics in Medicine and Biology. Lecture Notes in Biomathematics*, vol. 57. – Springer-Verlag, 1985.

[11] Waterman (M.). – *Introduction to Computational Biology.* – Chapman and Hall, London, 1995.

[12] Waterman (Michael S.). – *Secondary structure of single-stranded nucleic acids*, pp. 167–212. – Academic Press, 1978, *Advances in Mathematics Supplementary Studies.*

[13] Waterman (Michael S.). – Combinatorics of RNA hairpins and cloverleaves. *Studies in Applied Mathematics*, vol. 60, n° 2, 1979, pp. 91–96.

# Coverage Processes in Physical Mapping by Anchoring Random Clones

*Sophie Schbath*
INRA

February 10, 1997

[summary by Mireille Régnier]

## 1. Introduction

A complete physical map of the DNA of an organism consists of ordered overlapping fragments spanning the entire genome. A large number of fragments, called *clones*, are chosen at random from a library in which the entire genome is represented. The anchoring approach, efficient for large clones and then for very large genomes, uses an additional random genomic library of *anchors* and is based on the anchor-content of the clones. These anchors consist of any short sequence of DNA that occurs exactly once in the genome. Thus, clones containing an anchor in common overlap. Clones that overlap are then assembled into *islands* which cover some regions of the genome.

To plan a physical mapping project, it is therefore important to study, for instance, the proportion of the genome covered by islands, the number and the length of islands, with respect to the number of clones and anchors considered. To study the statistical properties of the islands, we must model the clones and anchors processes. This problem is typically part of the theory of coverage processes [6]. The aim of this paper is to provide a mathematical analysis of physical mapping by anchoring random clones, in a general model.

## 2. Notation and preliminary results

We consider the genome linearly as a discrete sequence of bases of length $G$. We assume that clones have independent and identically distributed lengths $L$, with mean $\mathcal{E}L$. We define $g = G/\mathcal{E}L$. Since the anchor length is very small compared to the clone length, the anchors will be considered to be points. We assume that right ends of clones occur on the real line according to a non-homogeneous Poisson process of rate $\alpha(t)$, and we label them $\{C_i, i \in \mathbb{Z}\}$ such that $\cdots < C_{-1} < C_0 \leq 0 < C_1 < \cdots < C_{N(g)} \leq g < C_{N(g)+1} < \cdots$; $N(t) \equiv N((0,t])$ denotes the counting process of clones and represents the number of clones ending in $(0,t]$. Similarly, the anchors occur according to a non-homogeneous Poisson process of rate $\lambda(t)$ and are labelled $\{A_j, j \in \mathbb{Z}\}$ such that $\cdots < A_{-1} < A_0 \leq 0 < A_1 < \cdots < A_{M(g)} \leq g < A_{M(g)+1} < \cdots$; $M(t) \equiv M((0,t])$ denotes the counting process of anchors and represents the number of anchors in $(0,t]$. We assume the process of anchors and the process of clones are independent, and the rates $\alpha$ and $\lambda$ are positive functions such that their integrals are finite on bounded sets, but are not finite on unbounded sets.

PROPOSITION 1. *The probability $J(t;x)$ that the points $t$ and $t+x$ $(x>0)$ are not covered by a common clone is*

$$J(t;x) = \exp\left(-\int_{x_{19}}^{\infty} \alpha(t+u)\mathcal{F}(u)du\right).$$

## 3. Main results

We introduce the following terminology: a clone containing an anchor is an anchored clone, and the clones containing one or more common anchors are assembled into anchored islands An anchored island can be composed of a unique anchored clone. A region between two anchored islands is an ocean.

Several theorems give properties of unanchored clones, anchored islands and notably the singleton anchored islands, composed of a unique anchored clone. The next theorem is about the proportion of oceans, in other words the proportion of the genome not covered by anchored islands.

THEOREM 1. *The probability $r_0(t)$ that $t$ is not covered by any anchored island is*

$$r_0(t) = \int_0^\infty \int_0^\infty \frac{J(t;v)J(t-w;w)}{J(t-w;v+w)} \lambda(t+v)\lambda(t-w) \exp\left(-\int_{t-w}^{t+v} \lambda(x)dx\right) dv\, dw.$$

*The mean proportion $r_0$ of the genome not covered by anchored islands is*

$$r_0 = \frac{1}{g}\int_0^g r_0(t)\, dt.$$

Under probabilistic assumptions that are satisfied for periodic rates $\alpha$ and $\lambda$, the author states the weak law of large numbers for the number of anchored islands and the number of anchored clones. Moreover, the length of the anchored islands, the mean number of clones and the mean number of anchors in an anchored island are derived. Notably, we get:

THEOREM 2. *(i) The process of anchors covered by clones is a Poisson process with rate given by*

$$\nu(t) = \lambda(t)(1 - J(t;0));$$

*(ii) the average number of anchors per anchored island ending in $(0,g]$, denoted by $\overline{H}_g$, is such that*

$$\overline{H}_g - \frac{\int_0^g \lambda(t)(1-J(t;0))\, dt}{\int_0^g \alpha(t)p_1(t)\, dt} \xrightarrow{\Pr} 0 \quad as \quad g \to \infty.$$

*where $p_1(t)$ is the probability that a clone ending at $t$ is the rightmost clone of an anchored island.*

Whereas an ocean is a region not covered by anchored islands, an actual ocean is defined to be a region not covered by clones.

THEOREM 3. *The probability $p_3(t;x)$ that an anchored island ending at $t$ is followed by an actual ocean of length at least $x$ is*

$$p_3(t;x) = J(t+x;0) \exp\left(-\int_t^{t+x} \alpha(u)du\right) \frac{1-q_1(t)}{p_1(t)}.$$

*where $q_1(t)$ is the probability that a clone ending at $t$ contains no anchor.*

## 4. Applications

The author presents numerical results for a genome of length $G = 100,000$ kb, 2300 clones of fixed length $L = 250$ kb and 500 anchors, that corresponds approximately to the physical mapping project of *Arabidopsis thaliana* genome [5]. The normalized genome length is $g = 400$, and the mean number of clones and anchors in $(0, g]$ are respectively $\mathcal{E}N \equiv \mathcal{E}N(g) = 2300$ and $\mathcal{E}M \equiv \mathcal{E}M(g) = 500$.

We focus on four quantities, namely the mean number of unanchored clones, the mean number of anchored islands, the average length of anchored islands and the mean proportion of the genome not covered by anchored islands. The first non-homogeneous model assumes each rate is piecewise constant and can take two values. It means the genome is composed of alternating rich and poor regions of clones, and of anchors. This is the *hotspot* model introduced by [4]. The second model allows some *sinusoidal fluctuations* around the constant rates.

**4.1. Hotspot model.** These are the main trends. When the hotspots are in phase, the number of unanchored clones decreases as the level of clone hotspot increases. As both of the hotspot levels increase, the mean number of anchored islands and the mean length of an anchored island decrease, leading to an increase of the mean proportion of the genome not covered by anchored islands.

When the hotspots are completely out of phase, the number of unanchored clones increases as both of the hotspot levels increase. As the level of clone hotspot increases, the mean number of anchored islands increases and then drops, whereas it decreases as the level of anchor hotspot increases. The mean length of anchored islands decreases as the level of clone hotspot increases. The mean proportion of the genome not covered by anchored islands increases as the levels of hotspot increase, rather dramatically when the level of clone hotspot increases.

The most important effect is undeniably that the more the model departs from the stationary model, the greater the mean proportion of genome not covered by anchored islands. Finally, the simulation results given by [4] and the theoretical calculations are quite close.

**4.2. Sinusoidal model.** We consider the following rates

$$\alpha(t) = \alpha_0 + \alpha_1 \sin(\alpha_2 t), \qquad \lambda(t) = \lambda_0 + \lambda_1 \sin(\lambda_2 t).$$

The mean number of unanchored clones seems not to depend on $\alpha_1$ and increases as $\lambda_1$ increases. Increasing either $\alpha_1$ or $\lambda_1$ has the same effect on the average length of anchored islands and the mean proportion of genome not covered by anchored islands: the length decreases whereas the proportion increases. The stationary case, $\alpha_1 = \lambda_1 = 0$ minimizes the mean number of anchored islands and the mean proportion of genome not covered by anchored islands and maximizes the average length of anchored islands.

## 5. Conclusion

Undeniably, the inhomogeneity in the clone and anchor locations along the genome substantially changes the predictions in a physical mapping project. Since the goal of a physical mapping project (at least the first step) is to obtain few long anchored islands and a small proportion of genome not covered by anchored islands, the previous applications clearly show that using homogeneous Poisson processes for clones and anchors provides an overly optimistic assessment of the progress of the mapping project. The difficulty in practice remains to model the inhomogeneity occurring in the genome. The detection of regions rich or poor in restriction sites involved in the cloning process, for instance, could be a first step in characterizing some hotspots. Since longer DNA sequences are becoming available, modelling heterogeneity has become an important problem in sequence analysis.

## Bibliography


[1] Arratia (R.), Lander (E. S.), Tavaré (S.), and Waterman (M. S.). – Genomic mapping by anchoring random clones: A mathematical analysis. *Genomics*, vol. 11, 1991, pp. 806–827.

[2] Barillot (E.), Dausset (J.), and Cohen (D.). – Theoretical analysis of a physical mapping strategy using random single-copy landmarks. *Proceedings of the National Academy of Sciences of the USA*, vol. 88, 1991, pp. 3917–3921.

[3] Chumakov (I.), Rigault (P.), Guillou (S.), Ougen (P.), Billaut (A.), Guasconi (G.), Gervy (P.), Le Gall (I.), Soularue (P.), and Grinas (L.). – Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature*, n° 359, 1992, pp. 380–387.

[4] Ewens (W. J.). – *Simulation results for anchored clones.* – Research Report n° 252, Department of Mathematics, Monash University, Australia, 1996. 7 pages.

[5] Ewens (W. J.), Bell (C. J.), Donnelly (P. J.), Dunn (P.), Matallana (E.), and Ecker (J. R.). – Genome mapping with anchored clones: Theoretical aspects. *Genomics*, vol. 11, 1991, pp. 799–805.

[6] Hall (Peter). – *Introduction to the Theory of Coverage Processes.* – John Wiley & Sons Inc., New York, 1988, *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, xx+408p.

[7] Hudson (T.), Stein (L.), Gerety, Ma (J.), Castle (A.), Silva (J.), Slonim (D.), Baptista (R.), Kruglyak (L.), Xu (S.), Hu (X.), Colbert (A.), Rosenberg (C.), Reeve-Daly (M.), Rozen (S.), Hui (L.), Wu (X.), Vestergaard (C.), Wilson (K.), Bae (J.), Maitra (S.), Ganiatsas (S.), Evans (C.), DeAngelis (M.), Ingalls (K.), Nahf (R.), Horton (L.), Oskin Anderson (M.), Collymore (A.), Ye (W.), Kouyoumjian (V.), Zemsteva (I.), Tam (J.), Devine (R.), Courtney (D.), Renaud (M.), Nguyen (H.), O'Connor (T.), Fizames (C.), Fauré (S.), Gyapay (G.), Dib (C.), Morissette (J.), Orlin (J.), Birren (B.), Goodman (N.), Weissenbach (J.), Hawkins (T.), Foote (S.), Page (D.), and Lander (E. S.). – An STS-based map of the human genome. *Science*, vol. 270, 1995, pp. 1945–1954.

[8] Karlin (S.) and Macken (C.). – Some statistical problems in the assessment of inhomogeneities of DNA sequence data. *Journal of the American Statistical Association*, vol. 86, 1991, pp. 27–35.

[9] Lander (E. S.) and Waterman (M. S.). – Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics*, vol. 2, 1988, pp. 231–239.

[10] Lee (W.). – *A mathematical analysis of genome physical mapping.* – Master's thesis, Department of Mathematics, Unversity of Southern California, 1992.

[11] Marr (G. T.), Yan (X.), and Yu (Q.). – Genomic mapping by single copy landmark detection: A predictive model with a discrete mathematical approach. *Mamm. Genome*, vol. 3, 1992, pp. 644–649.

[12] Nelson (D. O.) and Speed (T. P.). – Predicting progress in directed mapping projects. *Genomics*, vol. 24, 1994, pp. 41–52.

[13] Olson (M. V.), Dutchik (J. E), Graham (M. Y.), Brodeur (G. M.), Helms (C.), Frank (M.), MacCollin (M.), Scheinman (R.), and Frank (T.). – Random-clone strategy for genomic restriction mapping in yeast. *Proceedings of the National Academy of Sciences of the USA*, vol. 83, 1986, pp. 7826–7830.

[14] Port (E.), Sun (F.), Martin (D.), and Waterman (M. S.). – Genomic mapping by end-characterized random clones: A mathematical analysis. *Genomics*, vol. 26, 1995, pp. 84–100.

[15] Torney (D. C.). – Mapping using unique sequences. *Journal of Molecular Biology*, vol. 217, 1991, pp. 259–264.

[16] Zhang (M. Q.) and Marr (T. G.). – Genome mapping by nonrandom anchoring: A discrete theoretical analysis. *Proceedings of the National Academy of Sciences of the USA*, vol. 90, 1991, pp. 600–604.

# Heaps of Coins: Performance Evaluation and Task Resource Models

*Jean Mairesse*
LIAFA, Université Paris 7

June 2, 1997

**Abstract**

We introduce a model which generalizes the representation of trace monoids as heaps of coins (Viennot 86). Intuitively, one stacks (à la "Tetris") coins with a polyomino shape. This model makes it possible to study the behaviour of 1-bounded temporized Petri nets. The height of a heap of coins is recognized by an automaton with multiplicities in the semi-ring (max,+).

Using this representation as a starting point, various results of performance analysis are obtained: asymptotic average-case, worst-case and best-case behaviours, optimal scheduling.

# Part 2

# Symbolic Computation

# New Algorithms for Definite Summation and Integration

*Frédéric Chyzak*
Projet Algo – Inria Rocquencourt

March 17, 1997

[summary by Bruno Salvy]

**Abstract**

In 1978, W. Gosper gave an algorithm to compute the indefinite sum of an hypergeometric sequence. This algorithm has been incorporated in most computer algebra systems as the basis of their summation routines. Then, in the early 1990's D. Zeilberger applied a version of Gosper's algorithm in a clever way to the efficient calculation of definite sums of hypergeometric sequences. Zeilberger also gave a very general but slow algorithm for the general case of holonomic functions. F. Chyzak shows how Zeilberger's fast algorithm can be extended to a much more general context, including summation and integration of holonomic functions and sequences.

## Introduction

F. Chyzak's algorithms [7] aim at computing automatically the right-hand side of identities like the following ones, given their left-hand side:

$$\sum_{k=1}^{n} \binom{k}{m} H_k = \binom{n+1}{m+1}\left(H_{n+1} - \frac{1}{m+1}\right),$$

$$\sum_{k=0}^{n}\left(\sum_{j=0}^{k}\binom{n}{k}\right)^3 = n2^{3n-1} + 2^{3n} - 3n2^{n-2}\binom{2n}{n},$$

$$\sum_{n=0}^{\infty} H_n(x)H_n(y)\frac{u^n}{n!} = \frac{\exp\left(\frac{4u(xy-u(x^2+y^2))}{1-4u^2}\right)}{\sqrt{1-u^2}},$$

$$\int_{-1}^{+1} \frac{e^{-px}T_n(x)}{\sqrt{1-x^2}}dx = (-1)^n \pi I_n(p),$$

$$\int_0^{+\infty} xe^{-px^2} J_n(bx)I_n(cx)dx = \frac{1}{2p}\exp\left(\frac{c^2-b^2}{4p}\right) J_n\left(\frac{bc}{2p}\right),$$

$$\int_0^{+\infty} xJ_1(ax)I_1(ax)Y_0(x)K_0(x)dx = -\frac{\ln(1-a^4)}{2\pi a^2},$$

$$\sum_{n=0}^{\infty} \frac{r_n(a,b)z^n}{(q;q)_n} = \frac{1}{(az;q)_\infty (bz;q)_\infty}.$$

More precisely, for indefinite summations or integrations, the output of the algorithm is an expression for the right-hand side in terms of the left-hand side, while in the definite case, these new algorithms will produce a linear operator or a system of linear operators annihilating the right-hand side. From there, the solution can be found by Petkovšek's algorithm [9] in the recurrence case, by its $q$-version [4, 5] in the $q$-case, or by algorithms on differential equations [10, 11] in the differential case.

These examples are treated in a unified manner by considering algebras of linear operators $\partial_1, \ldots, \partial_r$ with coefficients in a suitable field $\mathbb{K}$. In most applications, these operators are partial differentiations, shifts or $q$-shifts. Then one considers *$\partial$-finite terms* $t$ which are such that the

$$\boldsymbol{\partial^\alpha} \cdot t = \partial_1^{\alpha_1} \cdots \partial_r^{\alpha_r} \cdot t, \qquad \alpha_i \in \mathbb{N}, \tag{1}$$

span a finite-dimensional vector space over $\mathbb{K}$.
Examples are:

- $\exp(z^2)$ for $D_z$ (derivation) over $\mathbb{Q}(z)$ (with dimension 1);
- the binomial coefficient $\binom{n}{k}$ for $S_n$ and $S_k$ (shifts) over $\mathbb{Q}(n,k)$ (with dimension 1);
- the Bessel $J_n(z)$ functions, the Tchebychev polynomials $T_n(z)$, the Legendre polynomials $P_n(z)$ for $S_n$ (shift) and $D_z$ (derivation) over $\mathbb{Q}(n,z)$ (with dimension 2);
- the product of Bessel functions $J_1(ax)I_1(ax)Y_0(x)K_0(x)$ for $D_a$ and $D_x$ (derivations) over $\mathbb{Q}(a,x)$ (with dimension 16).

In the case of shifts, $\partial$-finite terms corresponding to 1-dimensional vector spaces are exactly the hypergeometric sequences. In the differential case, Zeilberger's holonomic functions [12] are also $\partial$-finite. The closure properties under sum, product and the $\partial_i$'s generalize to this context [8].

## 1. Indefinite $\partial$-finite summation and integration

Given a $\partial$-finite term and a basis $b_1, \ldots, b_N$ of the vector space generated by its $\boldsymbol{\partial^\alpha}$ as in (1), Chyzak's first algorithm finds solutions in this vector space for equations of the form

$$P(\partial_1, \ldots, \partial_r) \cdot X = B,$$

where $P$ is a polynomial in $\mathbb{K}\langle \partial_1, \ldots, \partial_r \rangle$, $X$ is the unknown and $B$ is a given element of the vector space. The first step of the algorithm consists in setting

$$X = \phi_1 b_1 + \cdots + \phi_N b_N$$

with undeterminate coefficients $\phi_i$'s in $\mathbb{K}$. The left-hand side of the equation is then expressed in the basis of the $b_i$'s. Identifying coefficients then yields a system of linear equations satisfied by the $\phi_i$'s. It then suffices to find *rational* solutions of this system, i.e., solutions in $\mathbb{K}$. Several algorithms developed by S. Abramov and coauthors are available for this purpose, depending on the $\partial_i$'s [1, 2, 3, 6].

An important special case of this algorithm is when $P = S_n - 1$, which correspond to indefinite summation. In this setting the problem solved by Gosper's algorithm corresponds to vector spaces of dimension 1, and then a single rational coefficient has to be found.

For instance, consider computing the primitive of the integral cosine,

$$\mathrm{Ci}(z) = \int_0^z \frac{\cos(t)}{t}\, dt.$$

From the differential equation satisfied by $\cos(t)$, one readily computes a third order linear differential equation satisfied by Ci. In the algebra $\mathcal{A} = \mathbb{Q}(z)\langle D_z \rangle$, we thus work in the vector space

generated by $\mathrm{Ci}, \mathrm{Ci}', \mathrm{Ci}''$. Thus we look for

$$T = \phi_0 \,\mathrm{Ci} + \phi_1 \,\mathrm{Ci}' + \phi_2 \,\mathrm{Ci}'', \qquad \text{such that} \qquad D_z \cdot T = \mathrm{Ci}\,.$$

This leads to a simple system whose only rational solution is

$$\phi_0(z) = z, \qquad \phi_1(z) = 1, \qquad \phi_2(z) = z,$$

or in other words

$$\int \mathrm{Ci}(z)\, dz = z\,\mathrm{Ci}(z) - \sin(z).$$

## 2. Definite $\partial$-finite summation and integration

Zeilberger's algorithms for definite summation and integration are based on *creative telescoping*. To simplify, we describe this method in the case of summation, i.e. we consider a $\partial$-finite term $t_n$ when $\partial_1 = S_n$ is a shift with respect to a variable $n$, and we want to compute

$$\sum_{n=a}^{b} t_n,$$

or more precisely an operator annihilating this sum.

If we can find two polynomials $P$ and $Q$ in the algebra such that $P$ commutes with $\sum_n$ and

$$P \cdot t_n = (S_n - 1) Q \cdot t_n, \tag{2}$$

then interchanging $P$ and $\sum$ we get

$$P \cdot \sum_{n=1}^{b} t_n = [Q \cdot t_n]_a^b.$$

When moreover the bounds $a$ and $b$ are such that $t_n$ and all its $\boldsymbol{\partial^\alpha}$'s are zero there (the so-called case of *natural* boundaries), then this computation has for consequence that the right-hand side telescopes, whence the name of the method.

In the hypergeometric case, Zeilberger's fast algorithm [13] to find $P$ and $Q$ relies on the observation that (2) is equivalent to $P \cdot t_n$ being Gosper-summable. He then shows that Gosper's algorithm can be extended to handle undeterminate coefficients in $P$ and find conditions on these coefficients for the sum to be hypergeometric. The algorithm then increases the order of $P$ (and consequently the number of indeterminate coefficients) till the extended Gosper algorithm finds a sum. Termination is guaranteed via Bernstein's theory of holonomic functions.

Chyzak's second algorithm is an extension of Zeilberger's algorithm to the more general context of $\partial$-finite terms. Gosper's algorithm is replaced by the algorithm of the previous section, which is first extended to handle undeterminate coefficients and find conditions on these coefficients for an indefinite $\partial^{-1}$ to exist in the vector space. Termination is guaranteed only when Bernstein's theory can be invoked.

As an example, consider Neumann's addition theorem on Bessel functions:

$$J_0(z)^2 + 2\sum_{k=1}^{\infty} J_k(z)^2 = 1. \tag{3}$$

In the algebra $\mathcal{A} = \mathbb{Q}(z,k)\langle D_z, S_k\rangle$, the sequence of functions $J_k(z)^2$ satisfies the system

$$\begin{cases} zD_z^2 + (-2k+1)D_z - 2S_k z + 2z, \\ zD_z S_k + zD_z + (2k+2)S_k - 2k, \\ z^2 S_k^2 - 4(k+1)^2 S_k - 2z(k+1)D_z + 4k(k+1) - z^2, \end{cases}$$

from which follows that $J_k(z)^2$ is $\partial$-finite and generates a vector space of dimension 3 with basis $\{J_k(z)^2, S_k \cdot J_k(z)^2, D_z \cdot J_k(z)^2\}$. The output of the algorithm is

$$P = D_z, \qquad Q = \frac{k}{z} + \frac{1}{2} D_z,$$

which means that

$$D_z \cdot \sum_{k=0}^{\infty} J_k(z)^2 + \left[Q \cdot J_k(z)^2\right]_{k=0}^{\infty} = 0.$$

After some rewriting and considering the initial conditions, this is indeed equivalent to (3).

## Bibliography


[1] Abramov (S. A.). – Rational solutions of linear differential and difference equations with polynomial coefficients. *USSR Computational Mathematics and Mathematical Physics*, vol. 29, n° 11, 1989, pp. 1611–1620. – Translation of the Zhurnal vychislitel'noi matematiki i matematichesckoi fiziki.

[2] Abramov (S. A.). – Rational solutions of linear difference and $q$-difference equations with polynomial coefficients. In Levelt (A.) (editor), *Symbolic and algebraic computation*. pp. 285–289. – New York, 1995. Proceedings of ISSAC'95, Montreal.

[3] Abramov (S. A.) and Kvashenko (K. Yu.). – Fast algorithms to search for the rational solutions of linear differential equations with polynomial coefficients. In Watt (Stephen M.) (editor), *Symbolic and algebraic computation*. pp. 267–270. – New York, 1991. Proceedings of ISSAC'91, Bonn.

[4] Abramov (Sergei A.), Paule (Peter), and Petkovšek (Marko). – $q$-hypergeometric solutions of $q$-difference equations, 1996.

[5] Abramov (Sergei A.) and Petkovšek (Marko). – Finding all $q$-hypergeometric solutions of $q$-difference equations. In Leclerc (B.) and Thibon (J. Y.) (editors), *Formal power series and algebraic combinatorics*. pp. 1–10. – Université de Marne-la-Vallée, 1995. Proceedings SFCA'95.

[6] Abramov (Sergei A.) and Zima (Eugene V.). – A universal program to uncouple linear systems, 1996. Preprint.

[7] Chyzak (Frédéric). – An extension of Zeilberger's fast algorithm to general holonomic functions. In *Formal Power Series and Algebraic Combinatorics*. pp. 172–183. – Wien, Austria, 1997. Also available as INRIA Research Report 3195.

[8] Chyzak (Frédéric) and Salvy (Bruno). – *Non-commutative Elimination in Ore Algebras Proves Multivariate Holonomic Identities.* – Research Report n° 2799, Institut National de Recherche en Informatique et en Automatique, February 1996.

[9] Petkovšek (Marko). – Hypergeometric solutions of linear recurrences with polynomial coefficients. *Journal of Symbolic Computation*, vol. 14, 1992, pp. 243–264.

[10] Petkovšek (Marko) and Salvy (Bruno). – Finding all hypergeometric solutions of linear differential equations. In Bronstein (Manuel) (editor), *ISSAC'93*. pp. 27–33. – New York, July 1993.

[11] Singer (Michael F.). – Formal solutions of differential equations. *Journal of Symbolic Computation*, vol. 10, 1990, pp. 59–94.

[12] Zeilberger (Doron). – A holonomic systems approach to special functions identities. *Journal of Computational and Applied Mathematics*, vol. 32, n° 3, 1990, pp. 321–368.

[13] Zeilberger (Doron). – The method of creative telescoping. *Journal of Symbolic Computation*, vol. 11, 1991, pp. 195–204.

# An Efficient Algorithm to Compute the Rational Solutions of a Linear Differential System

*Moulay Barkatou*
LMC-IMAG Université de Grenoble I

January 27, 1997

[summary by Philippe Dumas]

The talk is in two parts: first a presentation of Abramov's algorithm for linear differential equations, next a generalisation to differential systems.

The aim of Abramov's algorithm [3] is to find all rational solutions of a linear differential equation with polynomial coefficients, that is an equation of the form

$$a_n(x)y^{(n)}(x) + \cdots + a_1(x)y'(x) + a_0(x)y(x) = b(x), \tag{1}$$

where $a_n(x), \ldots, a_0(x)$, $b(x)$ are polynomials. The first step of the algorithm is to bound the denominator $g(x)$ of a possible solution $f(x)/g(x)$. By a bound, we mean a polynomial $q(x)$ such that $g(x)$ divides $q(x)$. The coefficients of the polynomials $a_k(x)$ lie in a number field $\mathbb{K}$. The key argument uses the splitting field $\widehat{\mathbb{K}}$ of the polynomial $a_n(x)$. If $\xi$ is a root of $a_n(x)$ and the coefficients read

$$a_k(x) = (x-\xi)^{\beta_k} h_k(x), \qquad h_k(\xi) \neq 0, \quad k = 0, \ldots, n,$$

a rational solution $y(x)$, which necessarily belongs to $\widehat{\mathbb{K}}(x)$, may be written

$$y(x) = (x-\xi)^r z(x)$$

where $\xi$ is neither a zero nor a pole of $z(x)$. The term $a_k(x)y^{(k)}(x)$ of (1) behaves like

$$r(r-1)\cdots(r-k+1)\frac{a_k^{(\beta_k)}}{\beta_k!}(x-\xi)^{r+\beta_k-k}$$

near $\xi$. But the right-hand side of (1) is a polynomial, hence the exponent $r$ is larger or equal to the minimal integer root $-\rho$ of the *indicial equation* $E_\xi(r)$; this equation is the coefficient of the lowest power of $x-\xi$ in

$$\sum_{k=0}^{n} r(r-1)\cdots(r-k+1)\frac{a_k^{(\beta_k)}}{\beta_k!}(x-\xi)^{\beta_k-k}.$$

This crude description does not give an efficient way to obtain a bound of the denominator $g(x)$. Indeed it is possible to avoid any use of $\widehat{\mathbb{K}}(x)$ by greatest common divisor techniques. Essentially the previous argumentation is applied to each irreducible factor $f(x)$ of $a_n(x)$ in place of each $x-\xi$ to give an indicial equation $E_f(r) = 0$. As a result, a bound $q(x)$ for the denominator is obtained.

Next $p(x)/q(x)$ substitutes for $y(x)$ into (1) and the polynomial $p(x)$ is to be determined. This is the second step of the algorithm. The degree of $p(x)$ is the order of $\infty$ as a pole of a solution, and the technique of the first step provides a bound $N$ for the degree of a polynomial solution. At this point it is possible to proceed by brute force collection of like powers of $x$, but the bound $N$ is not accurate.

It is more efficient to find the coefficient $p_N$ of $x^N$ in $p(x)$ and to write $p(x) = p_N x^N + p^*(x)$ with $p^*(x)$ a polynomial of degree at most $N-1$. In this manner, the coefficients of $p(x)$ are obtained in turn. If there is no rational solution the process stops at some stage.

The classical approach for the search of rational solutions of linear differential systems consists in reducing them to the scalar case. This idea goes back to Birkhoff [5, 7]. The reduction uses the concept of cyclic vectors. Its major drawback is the growth of coefficients.

Barkatou [4] proposes a direct approach to the linear differential systems, which uses the same ideas as Abramov's algorithm. The difference is the use of matrices instead of scalars, as exemplified in [6] or [8]. The differential system under consideration reads

$$d(x)Y'(x) - A(x)Y(x) = B(x), \tag{2}$$

where $A(x)$ is a square matrix and $B(x)$ is a vector, both with polynomial components. The coefficient $d(x)$ is a polynomial; for each irreducible factor $f(x)$ of $d(x)$ there is an indicial equation $E_f(r)$. Contrary to the scalar case, the indicial equation may reduce to $0 = 0$, but a convenient linear change of variables eliminates this problem. The indicial equation gives a bound for each irreducible factor of $d(x)$, and a bound for a common denominator of the components of $Y(x)$. It remains to determine a polynomial solution of linear differential system, as in the scalar case.

## Bibliography


[1] Abramov (S. A.). – Rational solutions of linear differential and difference equations with polynomial coefficients. *USSR Computational Mathematics and Mathematical Physics*, vol. 29, n° 11, 1989, pp. 1611–1620. – Translation of the Zhurnal vychislitel'noi matematiki i matematichesckoi fiziki.

[2] Abramov (S. A.). – Rational solutions of linear difference and $q$-difference equations with polynomial coefficients. In Levelt (A.) (editor), *Symbolic and algebraic computation*. pp. 285–289. – New York, 1995. Proceedings of ISSAC'95, Montreal.

[3] Abramov (S. A.) and Kvashenko (K. Yu.). – Fast algorithms to search for the rational solutions of linear differential equations with polynomial coefficients. In Watt (Stephen M.) (editor), *Symbolic and algebraic computation*. pp. 267–270. – New York, 1991. Proceedings of ISSAC'91, Bonn.

[4] Barkatou (Moulay). – An efficient algorithm to compute the rational solutions of systems of linear differential equations. – Preprint, 1997.

[5] Birkhoff (G. D.). – Formal theory of irregular linear difference equations. *Acta Mathematica*, vol. 54, 1930, pp. 205–246.

[6] Coddington (E. A.) and Levinson (M.). – *Theory of Ordinary Differential Equations*. – McGraw-Hill, 1955. Twelfth reprint 1991 of the Tata McGraw-Hill edition 1972.

[7] Hilali (Abdelaziz). – *Solutions formelles de systèmes différentiels linéaires au voisinage d'un point singulier*. – Doctorat d'État, Université scientifique, technologique et médicale de Grenoble, June 1987.

[8] Wasow (Wolfgang). – *Asymptotic expansions for ordinary differential equations*. – Dover Publications Inc., New York, 1987, x+374p. Reprint of the John Wiley 1976 edition.

# Absolute Factorization of Differential Operators

*Jacques-Arthur Weil*
Université de Limoges

January 27, 1997

[summary by Frédéric Chyzak]

## 1. The Problem

Consider the linear ODE $y^{(n)}(x)+a_{n-1}(x)y^{(n-1)}(x)+\cdots+a_0(x)y(x)=0$, where the coefficients $a_i$ are rational functions of $k=C(x)$ for an algebraic closure $C$ of the rational number field $\mathbb{Q}$. Solving this equation is an easier task when the corresponding linear differential operator in $\partial = d/dx$,

$$L = \partial^n + a_{n-1}(x)\partial^{n-1} + \cdots + a_0(x),$$

admits a factorization $L = L_2L_1$ where the product denotes composition. The Leibniz rule

$$\partial \cdot ay = (ay)' = a'y + ay' = (a\partial + a') \cdot y \qquad (a \in k)$$

defines a degree on the non-commutative ring $\mathbb{A} = k[\partial]$, which makes it left and right Euclidean.

Consider the operator

$$L = \partial^4 - \frac{1}{4}\partial^3 + \frac{3}{4x^2}\partial^2 - x.$$

It can be proved to be *irreducible* in $\mathbb{A}$, i.e., it admits no factorization $L_2L_1$ in $\mathbb{A}$. However, $L$ factorizes over the extension ring $k(\sqrt{x})[\partial]$:

$$L = \left(\partial^2 - \frac{1}{x}\partial + \frac{3}{4x^2} - \sqrt{x}\right)\left(\partial^2 - \sqrt{x}\right) = \left(\partial^2 - \frac{1}{x}\partial + \frac{3}{4x^2} + \sqrt{x}\right)\left(\partial^2 + \sqrt{x}\right).$$

Note that since $\sqrt{x}$ and $-\sqrt{x}$ are algebraically and differentially indiscernable, the conjugates of a right factor of $L$ are other right factors of $L$. In the example above, $L$ is the *least common left multiple* of both conjugate right factors.

More generally, an operator $L \in \mathbb{A}$ is called *absolutely reducible* when there exists an algebraic extension $k_{\text{ext}}$ of $k$ such that $L$ is reducible in $\mathbb{A}_{\text{ext}} = k_{\text{ext}}[\partial]$ (for a suitable extension of the action of $\partial$ on $k_{\text{ext}}$). For an absolutely reducible operator $L$ with a right factor $L_1 \in \mathbb{A}_{\text{ext}}$, let $\tilde{L}$ be the least common left multiple of the algebraic conjugates of $L_1$. As a simple result of differential Galois theory, $\tilde{L}$ is stable under the action of the differential Galois group of the extension $\mathbb{A}_{\text{ext}}$ over $\mathbb{A}$ (to be defined in the next section). This entails that $\tilde{L} \in \mathbb{A}$. Since $\tilde{L}$ divides $L$, we have that $L$ is irreducible but absolutely reducible in $\mathbb{A}$ if and only if $L$ is the least common left multiple of the conjugates of a right factor $L_1 \in \mathbb{A}_{\text{ext}}$.

The example above motivates the following problems, sorted by increasing complexity:

1. find an algorithm to *decide* absolute reducibility;
2. find an algorithm to *compute* a factorization on an algebraic extension;
3. find an algorithm to compute a factorization on an algebraic extension with *absolutely irreducible factors*.

The algorithms to solve these problems, reduce to solving ODE's for solutions in special classes. A solution $y$ such that $y \in k$ is called a *rational solution*, while a solution $y$ such that $y'/y \in k$ is called an *exponential solution*[1] and a solution $y$ such that $y'/y$ is algebraic over $k$ is called a *Liouvillian solution*. An early study on this topic dates back to Liouville [6, 7]. The first algorithm to solve for rational solutions was developed in [1]. It relies on the resolution for polynomial solutions, for which an optimized algorithm is presented in [2]. Next, algorithms for factorization as well as algorithms to solve for Liouvillian solutions rely on the resolution for rational or exponential solutions. Algorithms for factorization are given in [3, 4, 9, 12]. The first algorithm to solve for Liouvillian solutions of second-order ODE's is due to Kovacic [5] and was later elaborated in [11], again in the second-order case. A prototypical algorithm for higher-order equations is to be found in [8] and was highly improved on in [10] in the third order case.

In the remainder of this summary, we comment on an algorithm to solve the second problem.

## 2. Differential Galois Theory

In the suitable analytical framework, the solution space $V$ of the equation $L \cdot y = 0$ is the $C$-vector space generated by $n$ *linearly independent* solutions $y_i$. However, these solutions satisfy *algebraic differential relations*

$$P_i\left(y_1, y_1', \ldots, y_1^{(n-1)}, \ldots, y_n, y_n', \ldots, y_n^{(n-1)}\right) = 0$$

for polynomials $P_i$ in $n^2$ variables and with coefficients in $k$. As an example, any solution $y_1$ of the equation $y'' + y = 0$ satisfies an algebraic equation $y_1^2 + y_1'^2 = c \in C$. For a given $L$, we would like to describe the ideal $\mathfrak{I}$ generated by all algebraic differential relations. A description is obtained by *differential Galois theory*.

For a *differential field extension* $K$ of $k$, the group of automorphisms $\sigma$ of $K$ that induce the identity on $k$ and such that $\sigma(f') = \sigma(f)'$ for $f \in K$ is called the *differential Galois group* of $K$ over $k$ and is denoted $\mathrm{Gal}(K/k)$. Let $K$ be $k\left(y_1, \ldots, y_1^{(n-1)}, \ldots, y_n, \ldots, y_n^{(n-1)}\right)$, i.e., the smallest differential field extension of $k$ which contains the $y_i$'s and does not extend the field of constants $C$. This field is called the *Picard-Vessiot extension* of $L$. The group $\mathrm{Gal}(K/k)$ is called the *differential Galois group* of $L$ and denoted $\mathrm{Gal}_k(L)$. A computational representation of $G$ is obtained as follows. Assume $y$ to satisfy $L \cdot y = 0$, then for any automorphism $\sigma \in G$, $L \cdot \sigma(y) = \sigma(L \cdot y) = 0$. In other words, each automorphism moves a solution of $L$ to another solution. Consequently, $\sigma(y)$ is a linear $C$-combination of the $y_i$'s with coefficients that are independent from $y$. This yields a matrix representation of $G$. Thus $G$ is linear algebraic and the ideal $\mathfrak{I}$ is stable under the action of $G$.

We now proceed to introduce a lemma which is crucial to the algorithm discussed in the next section. Assume that $L$ admits a right factor $L_1$ with solution space $V_1 \subset V$. For any $v_1 \in V_1$ and any automorphism $\sigma \in G$, $L_1 \cdot \sigma(v_1) = \sigma(L_1 \cdot v_1) = 0$, so that $V_1$ is stable under $G$. We want to prove a converse property.

For an $r$-tuple $(v_1, \ldots, v_r) \in K^r$, the Wronskian $\mathrm{Wr}(v_1, \ldots, v_r)$ is classically defined as the matrix $\left[v_i^{(j)}\right]$. The corresponding determinant induces an application from $K^r$ to $K$. This application is an alternate $r$-linear form and satisfies

$$\sigma(\det(\mathrm{Wr}(v_1, \ldots, v_r))) = \det(\mathrm{Wr}(\sigma(v_1), \ldots, \sigma(v_r)))$$

for any $\sigma \in G$. Below, we more intrinsically use $r$-exterior products, i.e., formal alternate $r$-linear symbols $v_1 \wedge \cdots \wedge v_r$ that satisfy $\sigma(v_1 \wedge \cdots \wedge v_r) = \sigma(v_1) \wedge \cdots \wedge \sigma(v_r)$ for any $\sigma \in G$.

[1] Such a solution is also frequently referred to as a *hyperexponential solution*.

Let us assume $V_1$ to be a 2-dimensional $C$-vector subspace of $V$ with basis $(f_1, f_2)$ and stable under the action of $G$. More specifically, for each $\sigma \in G$ there exist $c_{i,j}^{(\sigma)} \in C \setminus \{0\}$ such that

$$\sigma(f_i) = c_{i,1}^{(\sigma)} f_1 + c_{i,2}^{(\sigma)} f_2.$$

Then in the exterior power $\Lambda^2(V_1)$ where $f_1 \wedge f_1 = f_2 \wedge f_2 = 0$,

$$\sigma(f_1 \wedge f_2) = \sigma(f_1) \wedge \sigma(f_2) = (c_{1,1}c_{2,2} - c_{1,2}c_{2,1})(f_1 \wedge f_2).$$

More generally, assume that $V_1$ is a $C$-subspace of $V$ stable under $G$ and with dimension $\dim V_1 = r < n = \dim V$. Then, the exterior $r$-power $\bigwedge^r(V_1)$ is a 1-dimensional vector space with basis $\omega = f_1 \wedge \cdots \wedge f_r$. For each $\sigma \in G$, there exists a non-zero $c_\sigma \in C$ such that $\sigma(\omega) = c_\sigma \omega$. In fact, $c_\sigma = \det \sigma$ when $\sigma$ is viewed as a $C$-linear automorphism of $V_1$. Now, for $y \in V$, write

$$L_1 \cdot y = \frac{\det(\mathrm{Wr}(y, f_1, \ldots, f_r))}{\det(\mathrm{Wr}(f_1, \ldots, f_r))}.$$

This makes $L_1$ a linear operator of order $r$. For any $\sigma \in G$,

$$\sigma(L_1 \cdot y) = \frac{\sigma(\det(\mathrm{Wr}(y, f_1, \ldots, f_r)))}{\sigma(\det(\mathrm{Wr}(f_1, \ldots, f_r)))} = \frac{c_\sigma \sigma(\det(\mathrm{Wr}(y, f_1, \ldots, f_r)))}{c_\sigma \sigma(\det(\mathrm{Wr}(f_1, \ldots, f_r)))} = L_1 \cdot y.$$

The coefficients of $L_1$ are therefore left fixed by all elements of $G$, and $L_1 \in k[\partial]$.

LEMMA 1. *An operator $L$ with solution space $V$ admits a right factor $L_1$ such that the solution space $V_1$ of $L_1$ is a subspace of $V$ if and only if there exists a non-zero proper subspace of $V$ which is stable under $G$.*

## 3. The Beke-Bronstein Algorithm

Wronskians relate the solutions of an ODE to its coefficients. In particular, the Wronskian $w = \det(\mathrm{Wr}(y_1, \ldots, y_n)) = \det\left[Y, Y', \ldots, Y^{(n-1)}\right]$ where $Y$ is the column vector of the $y_i$'s satisfies

$$\begin{aligned} w' &= \sum_{i=1}^{n-1} \det\left[Y, \ldots, Y^{(i-1)}, Y^{(i+1)}, Y^{(i+1)}, \ldots, Y^{(n-1)}\right] + \det\left[Y, \ldots, Y^{(n-2)}, Y^{(n)}\right] \\ &= -\sum_{i=0}^{n-1} a_i(x) \det\left[Y, \ldots, Y^{(n-2)}, Y^{(i)}\right] = -a_{n-1}(x) \det\left[Y, Y', \ldots, Y^{(n-1)}\right] = -a_{n-1}(x) w. \end{aligned}$$

In short $w' + a_{n-1}(x)w = 0$ (*Liouville relation*); the other coefficients of $L$ satisfy similar relations.

The algorithm developed and implemented by Bronstein after Beke's work and described in [4] makes use of Wronskians in the following way. To obtain a right factor of the operator $L$:

1. solve $L \cdot y = 0$ for exponential solutions; if solutions are found, they yield first-order right factors of $L$;
2. similarly, find first-order left-hand factors by the method of adjoint operators [4]; if solutions are found, they yield right factors of $L$ of order $n - 1$;
3. if no solution was found, look for right factors of order $r$ $(2 \leq r \leq n - 2)$ as follows:
   (a) build an equation whose solution space is spanned by all Wronskians of order $r$;
   (b) solve for exponential solutions;
   (c) test which solutions are Wronskians, i.e., *pure* exterior products, and obtain a right factor.

As a comparison, Singer's method, which was implemented by Van Hoeij, relies on solving for rational solutions only.

## 4. An Example

Again, consider the operator $L = \partial^4 - \frac{1}{4}\partial^3 + \frac{3}{4x^2}\partial^2 - x$. Both first steps of the algorithm above fail, so that the only possible factorizations are of the form $L = L_2 L_1$ with factors of order 2. Write $w = y_1 y_2' - y_1' y_2$ for any two solutions of $L$. By computing its first derivatives, reducing them by $L$ on the basis $\left(y_1^{(i)} y_2^{(j)}\right)_{i,j=0,\ldots,3}$, and looking for linear dependencies by Gaussian elimination, we obtain that $w$ is annihilated by

$$P = \partial^5 - \frac{5}{2x}\partial^4 + \frac{21}{4x^2}\partial^3 - \frac{69}{8x^3}\partial^2 + \frac{8x^5+15}{2x^4}\partial.$$

The only exponential solutions are the constants $\lambda \in C$. This entails that $L_1 = \partial^2 - \lambda\partial + r(x)$ for an algebraic function $r$. By identification, one finds

$$L_2 = \partial^2 + \left(\lambda - \frac{1}{x}\right)\partial + \left(\lambda^2 - \frac{\lambda}{x} + \frac{3}{4x^2} - r(x)\right),$$

where $r(x) = \frac{1}{4x^2}\left(2\lambda^2x^2 - \lambda x \pm \sqrt{4\lambda^4x^4 - 8\lambda^3x^3 + 13\lambda^2x^2 - 15\lambda x + 16x^5}\right)$. Realizing that $\lambda = 0$, we get $r(x) = \pm\sqrt{x}$ and the factorizations of the first section.

## Bibliography


[1] Abramov (S. A.). – Rational solutions of linear differential and difference equations with polynomial coefficients. *USSR Computational Mathematics and Mathematical Physics*, vol. 29, n° 11, 1989, pp. 1611–1620. – Translation of the Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki.

[2] Abramov (Sergei A.), Bronstein (Manuel), and Petkovšek (Marko). – On polynomial solutions of linear operator equations. In Levelt (A.) (editor), *Symbolic and algebraic computation*. pp. 290–296. – New York, 1995.

[3] Beke (E.). – Die Irreducibilität des homogenen linearen Differentialgleichungen. *Mathematische Annalen*, vol. 45, 1884, pp. 278–294.

[4] Bronstein (M.) and Petkovšek (M.). – On Ore rings, linear operators and factorisation. *Programmirovanie*, n° 1, 1994, pp. 27–44. – Also available as Research Report 200, Informatik, ETH Zürich.

[5] Kovacic (Jerald J.). – An algorithm for solving second order linear homogeneous differential equations. *Journal of Symbolic Computation*, vol. 2, 1986, pp. 3–43.

[6] Liouville (J.). – Premier mémoire sur la détermination des intégrales dont la valeur est algébrique. *Journal de l'École polytechnique*, n° 14, 1833, pp. 124–148.

[7] Liouville (J.). – Second mémoire sur la détermination des intégrales dont la valeur est algébrique. *Journal de l'École polytechnique*, n° 14, 1833, pp. 149–193.

[8] Singer (Michael F.). – Liouvillian solutions of $n$-th order homogeneous linear differential equations. *American Journal of Mathematics*, vol. 103, n° 4, 1981, pp. 661–682.

[9] Singer (Michael F.). – Testing reducibility of linear differential operators: A group theoretic perspective. *Applicable Algebra in Engineering, Communication and Computing*, vol. 7, n° 2, 1996, pp. 77–104.

[10] Singer (Michael F.) and Ulmer (Felix). – Necessary conditions for Liouvillian solutions of (third order) linear differential equations. *Applicable Algebra in Engineering, Communication and Computing*, vol. 6, n° 1, 1995, pp. 1–22.

[11] Ulmer (Felix) and Weil (Jacques-Arthur). – *Note on Kovacic's algorithm*. – Prépublication n° 94-13, Institut de recherche mathématique de Rennes, Université de Rennes 1, France, July 1994.

[12] Van Hoeij (Mark). – Formal solutions and factorization of differential operators with power series coefficients. *Journal of Symbolic Computation*, vol. 24, n° 1, July 1997, pp. 1–30.

# Minimal Decomposition and Computation of Differential Bases for an Algebraic Differential Equation

*Évelyne Hubert*
LMC, Grenoble

September 21, 1996

[summary by Bruno Salvy]

**Abstract**

Singular and general solutions of algebraic differential equations can be expressed in a differential algebra setting regardless of the existence of closed-form. Theoretical and algorithmic tools in this area are presented.

## 1. Types of solutions

Consider the following differential equation:

$$y'^3 - 4xyy' + 8y^2 = 0. \tag{1}$$

This equation admits three solutions of different types:

$$y(x) = a(x-a)^2, \qquad y(x) = 0, \qquad y(x) = \frac{4}{27}x^3. \tag{2}$$

The first one is the *general* solution and the two other ones are *singular* solutions. The solution $y(x) = 0$ is actually a special case of the general solution and is called a *particular* singular solution; the third one is an *essential* singular solution. As showed by Figure 1, both singular solutions appear as envelopes of the general solution. In general, this is true of essential singular solutions of first order equations.

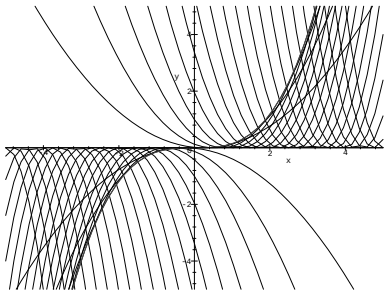


FIGURE 1. Solutions of (1)

In the general case of an equation

$$P(x, y, y', \ldots, y^{(n)}) = 0,$$

where $P$ is a polynomial, the *singular* solutions are the simultaneous solutions of $P$ and its *separant* $\partial P/\partial y^{(n)}$. Through differential algebra, a meaning can also be given to the "general" solution even when no closed-form exists. It is also possible to distinguish algebraically between a particular singular solution and an essential one. The aim of É. Hubert's work is to provide algorithms dealing with general and singular solutions in this framework.

To an ordinary differential equation like (1) is associated a differential polynomial

$$p = y_1^3 - 4xy_0y_1 + 8y_0^2 \in \mathbb{Q}(x)\{y\}.$$

More generally, the differential ring $\mathbb{A} = A\{y\}$ is a ring of polynomials in the variables $y_0, y_1, y_2, \ldots$ endowed with an operator $\delta$ which is a derivation on the commutative integral domain $A$ and which is such that $\delta y_i = y_{i+1}$. A *differential ideal* is an ideal stable under $\delta$. For instance the differential ideal generated by a differential polynomial $p$ is the polynomial ideal generated by $p$ and its derivatives:

$$[p] = (p, \delta p, \delta^2 p, \ldots).$$

Of particular interest is the *radical* of this ideal:

$$\{p\} = \sqrt{[p]} = \{a \in \mathbb{A} \mid \exists k \in \mathbb{N}, a^k \in [p]\}.$$

This is the set of differential polynomials vanishing on all the solutions of $p$. A differential ideal $I$ is *prime* when $ab \in I \Rightarrow a \in I$ or $b \in I$.

We now restrict to $A = \mathbb{Q}(x)$ for simplicity, but the results can be stated in much more generality, see [4, 5]. An important property is that a radical differential ideal $R$ can be decomposed into a finite intersection of prime differential ideals:

$$R = \bigcap_{k=1}^{r} P_k.$$

When none of the $P_k$ is included in another one, this decomposition is called *minimal* and is unique. These $P_k$'s are then called *essential components* of $R$. In the same way as $\{p\}$ corresponds to the solutions of $p$, each of the essential components of $\{p\}$ corresponds to one type of solutions of $p$. In the same example as before, a decomposition is

$$\{p\} = \{p, y_2^2 - 2xy_2 + 2y_1, y_3\} \cap \{y_0\} \cap \{27y_0 - 4x^3\},$$

each term corresponding to one of (2). This decomposition is not minimal since the second ideal obviously contains the first one and therefore corresponds to a particular singular solution. The minimal decomposition is obtained by removing this second term. Testing the inclusion of the general component (the first one here) into one of the other ones is related to *Ritt's problem*; its algorithmic resolution via the computation of a differential basis of each component is one of the aims of [3].

While the singular solutions obviously correspond to components of the ideal $\{p, s\}$, where $s$ is the separant of $p$, the general solution corresponds to the *quotient* of $\{p\}$ by $s$, where the quotient of a radical differential ideal $R$ by an element $s \in \mathbb{A}$ is defined as

$$R : s = \{a \in \mathbb{A} \mid sa \in R\},$$

which is itself a radical differential ideal. Two properties are of interest: for any non-empty subset $\sigma$ of $\mathbb{A}$, one has $\{\sigma\} = \{\sigma\} : s \cap \{\sigma, s\}$; when $p$ is irreducible as a polynomial in $y_0, y_1, \ldots$ and $s$ is its

separant, the ideal $\{p\} : s$ is prime. Thus $\mathcal{G}_p = \{p\} : s$ is an essential component of $\{p\}$ which is called the *general component* of $p$.

## 2. Algorithms for minimal decompositions

We now turn to the actual computation of the minimal decomposition

$$\{p\} = \mathcal{G}_p \cap R_1 \cap \cdots \cap R_k.$$

Ritt showed that each essential component $R_i$ is the general component of some differential polynomial $a_i$. The computation of a minimal decomposition then requires finding these $a_i$'s and insuring that the decomposition is minimal.

Ritt's *low power theorem* states that when an irreducible differential polynomial $p$ of a differential ring $A\{y_1, \ldots, y_n\}$ is contained in one of the ideals $\{y_i\}$, then $\{y_i\}$ is an essential component of $\{p\}$ if and only if the lowest degree terms of $p$ do not contain a derivative of $y_i$. This theorem can be used to give a criterion for a $\mathcal{G}_a$ to be an essential component of $\{p\}$ after a *preparation process* which rewrites $p$ as a differential polynomial in $a$:

$$s_a^k p = \sum_{\alpha_0, \ldots, \alpha_e} c_{\alpha_0, \ldots, \alpha_e} a_0^{\alpha_0} \cdots a_e^{\alpha_e},$$

where $s_a$ is the separant of $a$, $a_i = \delta^i a$, $c_{\alpha_0, \ldots, \alpha_e} \notin \mathcal{G}_a$ and $e$ is the difference between the order $n$ of $p$ and the order $m$ of $a$. This reduction process is performed in three stages: for $i$ from 1 to $e$, $a_i = \delta a_{i-1}$ can be rewritten $a_i = s_a y_{m+i} + t_i$ for some $t_i$. Then $y_n$ is replaced by $(a_e - t_e)/s_a$ in $p$. After normalization this yields

$$s_a^{k_e} p = \sum c_{\alpha_e} a_e^{\alpha_e},$$

where the coefficients do not involve derivatives higher than $y_{n-1}$. The process is then repeated with the $y_{n-i}$ in succession. This rewrites all the $y_{m+i}$ for $i > 0$. Then in each monomial $c_{\boldsymbol{\alpha}} a_1^{\alpha_1} \cdots a_e^{\alpha_e}$, the coefficient $c_{\boldsymbol{\alpha}}$ is rewritten $\tilde{c}_{\boldsymbol{\alpha}} a_0^{\alpha_0}$, where $\alpha_0$ is the largest integer $k$ such that $a^k$ divides $c_{\boldsymbol{\alpha}}$.

In our previous example, the reduction of $p$ in terms of $y_0$ is tautologous; the lowest degree terms of $p$ is $-4xy_0y_1 + 8y_0^2$ and thus $\{y_0\}$ is not an essential component. The reduction of $p$ in terms of $a = 27y_0 - 4x^3$ is more interesting. The separant $s_a = 27$ is constant and the first step of the reduction process yields $a_1 = 27y_1 - 12x^2$. Then $p$ is rewritten successively

$$\begin{aligned} p &= y_1^3 - 4xy_0y_1 + 8y_0^2, \\ 19683p &= a_1^3 + 36x^2a_1^2 - 108x(27y_0 - 4x^3)a_1 + 216(27y_0 - 2x^3)(27y_0 - 4x^3), \\ 19683p &= a_1^3 + 36x^3a_1^2 - 108xa_0a_1 + 216(27y_0 - 2x^3)a_0. \end{aligned}$$

The lowest degree term is $216(27y_0 - 2x^3)a_0$ which does not involve $a_1$ and thus $\{27y_0 - 4x^3\}$ is an essential component. (There also exists a simpler algorithm in this case since $p$ is of order 1 [2]).

The computation of the $a_i$'s relies on the Rosenfeld-Gröbner algorithm [1] which has been implemented in Maple by F. Boulier. Given a system $\Sigma$ of differential polynomials and a ranking, this algorithm computes a decomposition of the radical ideal $\{\Sigma\}$ as a finite intersection of radical differential ideals:

$$\{\Sigma\} = \mathcal{I}_1 \cap \cdots \cap \mathcal{I}_s,$$

each $\mathcal{I}_i$ being described by a system of polynomial equations and inequations and a characteristic set. This decomposition makes it possible to test membership in the $\mathcal{I}_i$'s and therefore in $\{\Sigma\}$ by simple reductions. Note that the $\mathcal{I}_i$'s are not necessarily prime.

A lemma of Lazard's combined with Ritt's result mentioned above shows that the $\mathcal{I}_i$'s corresponding to essential components of $\{p, s\}$ must have a characteristic set reduced to one differential polynomial. This makes it possible to filter out some of the radical ideals. Then prime differential ideals can be obtained by factorization. This leads to the following algorithm.

Input: An irreducible differential polynomial $p$.
Output: $a_1, \ldots, a_r$ such that $\mathcal{G}_p, \mathcal{G}_{a_1}, \ldots, \mathcal{G}_{a_r}$ are the essential components of $\{p\}$.
$G$:=Rosenfeld-Gröbner($[p, s]$);
$A := []$;
for each $R$ in $G$ with cardinality 1 do
    for each factor $b$ of $R[1]$ do
        if low-powers(preparation$(p, b)$)$=cb_0^\rho$ then $A := A \cup [b]$;
Return $A$.

In our example, the output of Rosenfeld-Gröbner is

$$\{y_0\}, \qquad \{27y_0 - 4x^3\},$$

from where the computations above have been performed.

It is possible to avoid the factorization and perform only gcd computations. In some cases, this algorithm can also be extended to compute a differential basis of $\mathcal{G}_p$. This is helpful to compute power series solutions when the initial conditions lie on a singular solution. We refer to [3] for details.

## Bibliography


[1] Boulier (F.), Lazard (D.), Ollivier (F.), and Petitot (M.). – Representation for the radical of a finitely generated differential ideal. In Levelt (A. H. M.) (editor), *Symbolic and Algebraic Computation*. pp. 158–166. – New York, 1995. Proceedings of ISSAC'95, July 1995, Montreal, Canada.

[2] Hubert (Évelyne). – The general solution of an ordinary differential equation. In Lakshman (Y. N.) (editor), *Symbolic and Algebraic Computation*. pp. 189–195. – New York, 1996. Proceedings ISSAC'96, Zürich, July 1996.

[3] Hubert (Évelyne). – *Étude algébrique et algorithmique des singularités des équations différentielles implicites*. – PhD thesis, Institut National Polytechnique de Grenoble, April 1997.

[4] Kolchin (E. R.). – *Differential algebra and algebraic groups*. – Academic Press, New York, 1973, *Pure and Applied Mathematics*, vol. 54, xviii+446p.

[5] Ritt (Joseph Fels). – *Differential Algebra*. – American Mathematical Society, New York, N. Y., 1950, *American Mathematical Society Colloquium Publications*, vol. XXXIII, viii+184p.

# Differential Equations, Nested Forms and Star Products

*John Shackell*
University of Kent at Canterbury, U.K.

September 9, 1997

[summary by Frédéric Chyzak]

**Abstract**

This presentation is in two parts. First, we recall the definition of two types of asymptotic expansions known as *nested form* and *nested expansions*. This theory makes it possible to adapt the asymptotic scale to the function under expansion and is based on the theory of Hardy fields [1]. Next, we suggest a reformulation of nested forms in terms of generalized products called *star products*, and a prospective theory of multivariate Hardy fields called *partial Hardy fields*.

## PART I: ASYMPTOTIC NESTED EXPANSIONS

### 1. Hardy fields

From the asymptotic viewpoint, $\mathcal{C}^\infty$ real-valued functions do not behave as nicely as holomorphic functions. In particular, asymptotic comparisons of functions that involve the symbols $O$, $o$ and $\sim$ cannot be termwise differentiated. A simple example is provided by $f(x) = x + \cos x \sim x$, whereas $f'(x)$ is not asymptotic to 1. In rough terms, this defect is due to allowing the functions to oscillate. A remedy to this problem is the use of *Hardy fields* instead of rings of functions. A construction is as follows. In order to deal with fields of functions, one first considers the ring $\mathcal{G}$ of *germs* of $\mathcal{C}^\infty$ real-valued functions at $+\infty$, where two functions are identified when they agree on a neighbourhood of $+\infty$. Then, a *Hardy field* $\mathcal{H}$ is a subring of $\mathcal{G}$ which is also a differential field. An example is $\mathbb{R}(x)$ viewed as a ring of germs with the usual derivation.

Considering fields of germs has nice consequences. A non-zero function $f$ of a Hardy field $\mathcal{H}$ is invertible with $\mathcal{C}^\infty$ inverse, so that asymptotically it never vanishes, and is therefore of asymptotically constant sign. This defines a total order on $\mathcal{H}$ by $f < g$ if and only if $g - f$ has asymptotically positive sign. The derivative $f'$ is also of asymptotically constant sign, so that $f$ is monotonic and tends to a limit in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ when $x$ goes to $+\infty$. Applying this property to the ratio $f/g$ of two functions in $\mathcal{H}$, we obtain that either $f = o(g)$ or $f \sim cg$ for a non-zero real constant $c$ or $g = o(f)$.

For $f$ in a Hardy field $\mathcal{H}$, an elementary result is that both differential ring extensions $\mathcal{H}(\exp f)$ and $\mathcal{H}(\ln f)$ are again Hardy fields. Let $\ell_0(x) = x$ and for $n \geq 0$, $\ell_{n+1}(x) = \ln|\ell_n(x)|$. Similarly, let $e_0(x) = x$ and for $n \geq 0$, $e_{n+1}(x) = \exp e_n(x)$. For $n < 0$, define $e_n = \ell_{-n}$ and $\ell_n = e_{-n}$. Then, for a Hardy field $\mathcal{H}$, there is a Hardy field extension containing each of the $\ell_n(f)$ and $e_n(f)$. Again for a function $f$ in a Hardy field $\mathcal{H}$, it follows from a theorem by Rosenlicht that there exists a smallest Hardy field containing $\mathbb{R}$, $f$, and each $g^d$ for $g > 0$ and $d \in \mathbb{R}$. This field is denoted $\mathbb{R}\langle\langle f \rangle\rangle$. A computationally interesting fact is that $f^\Delta = (\ln|f|)' = f'/f \in \mathbb{R}\langle\langle f \rangle\rangle$ even when $\ln f \notin \mathbb{R}\langle\langle f \rangle\rangle$.

## 2. Comparability classes

In view of various types of expansions, we need to extend asymptotic equivalence $\sim$ to coarser and coarser equivalences. In particular, beside asymptotic equivalence ($x$ is equivalent to $x+\ln x$ but not to $2x$), we need to consider asymptotic equivalence up to a non-zero constant factor ($x$ is equivalent to $2x$ but not to $x^2$), asymptotic equivalence up to powers ($x$ is equivalent to $x^2$ but not to $\exp x$). For two functions $f$ and $g$ in a Hardy field $\mathcal{H}$ and with limiting values 0, $-\infty$ or $+\infty$, we define $f \approx_n g$ to mean that there exists a non-zero constant $c \in \mathbb{R}$ such that $\ell_n(f) \sim c\ell_n(g)$. Furthermore, we agree that $f \approx_n g$ when both functions tend to finite non-zero limits. For each $n \geq 0$, this defines an equivalence relation on $\mathcal{H} \setminus \{0\}$. Call $\gamma_n(f)$ the class of $f \neq 0$ and $\Upsilon_n(\mathcal{H})$ the set of equivalence classes.

Beside this, we need to measure the accuracy of an expansion (a series in $\ln x$ is finer than a series in $x$). This is done by defining an order between equivalence classes. Set $\gamma_n(f) < \gamma_n(g)$ to mean $\ell_n(f) = o(\ell_n(g))$. For each $n$, this defines a total order on $\Upsilon_n(\mathcal{H})$: $\ell_n(f)/\ell_n(g)$ is in a suitable Hardy field extension of $\mathcal{H}$, so that it has a limit $a$ in $\overline{\mathbb{R}}$ (independent of the extension). Either $a = 0$ and $\gamma_n(f) < \gamma_n(g)$; or $a = \pm\infty$ and $\gamma_n(g) < \gamma_n(f)$; or $\gamma_n(f) = \gamma_n(g)$.

Here are a few examples:

$$\gamma_0(\ln x) < \gamma_0\left(\exp\left(\ell_2(x)^2\right)\right) < \gamma_0(x) = \gamma_0(x+\ln x) < \gamma_0\left(x^2\right) < \gamma_0\left(\exp\left(\ln(x)^2\right)\right) < \gamma_0(\exp(x)),$$

$$\gamma_1(\ln x) < \gamma_1\left(\exp\left(\ell_2(x)^2\right)\right) < \gamma_1(x) = \gamma_1(x+\ln x) = \gamma_1\left(x^2\right) < \gamma_1\left(\exp\left(\ln(x)^2\right)\right) < \gamma_1(\exp(x)),$$

$$\gamma_2(\ln x) = \gamma_2\left(\exp\left(\ell_2(x)^2\right)\right) < \gamma_2(x) = \gamma_2(x+\ln x) = \gamma_2\left(x^2\right) = \gamma_1\left(\exp\left(\ln(x)^2\right)\right) < \gamma_2(\exp(x)).$$

The previous equivalence relations extend known cases: $\gamma_0$ is the valuation map with the usual ordering reversed; the elements of $\Upsilon_1(\mathcal{H})$ are Rosenlicht's comparability classes.

## 3. From $\gamma_n$ to $\gamma_{n+1}$

In view of the examples above, one proves that the $\sim_n$ are coarser and coarser relations: if $\gamma_n(f) = \gamma_n(g)$, then $\gamma_{n+1}(f) = \gamma_{n+1}(g)$. Thus for $n \geq 0$, the map $\gamma_{n+1}$ factors through $\Upsilon_n(\mathcal{H})$: $\gamma_{n+1} = \eta_n \circ \gamma_n$ for a surjection $\eta_n$ from $\Upsilon_n(\mathcal{H})$ to $\Upsilon_{n+1}(\mathcal{H})$. Taking the direct limit, we get $\Upsilon_\infty(\mathcal{H})$ and $\gamma_\infty$ such that for any $k \geq 0$ and any non-zero $d \in \mathbb{R}$,

$$\gamma_\infty(\ln x) < \gamma_\infty(x) = \gamma_\infty\left(x^d\right) = \gamma_\infty\left(e_k\left(\ell_k^d(x)\right)\right) < \gamma_\infty(\exp x).$$

On the contrary, inequalities are not always preserved by the $\eta_n$'s: when $\gamma_n(f) < \gamma_n(g)$, i.e., $\ell_n(f) = o(\ell_n(g))$, it is not always true that $\gamma_{n+1}(f) \leq \gamma_{n+1}(g)$. For instance,

$$\gamma_0\left(x^{-1}\right) < \gamma_0(\ln x) < \gamma_0(x) \qquad \text{but} \qquad \gamma_1(\ln x) < \gamma_1(x) = \gamma_1\left(x^{-1}\right).$$

However, the property is valid when comparing functions that are infinite at $+\infty$, as shown by the examples of the previous section.

On the other hand, comparing $\gamma_2(f)$ to $\gamma_2(\ell_{p-1}(x))$ yields information on $\gamma_1\left(L_p(x)f^\Delta(x)\right)$ where $L_p(x) = 1/\ell_p'(x) = x\ell_1(x)\cdots\ell_{p-1}(x)$. The key idea is to restrict to functions $f$ with infinite or zero limit at $+\infty$ and to use *L'Hôpital's rule*: if $f, g \to 0$ or $\pm\infty$, $\frac{f(x)}{g(x)} \sim \frac{f'(x)}{g'(x)}$. Applying this rule to $\ell_2(f)$ and $\ell_{p+1}(x)$, we obtain

$$\frac{\ell_2(f)}{\ell_{p+1}(x)} \sim \frac{f^\Delta}{\ln|f|}\frac{\ell_p}{\ell_p'}, \qquad \text{and hence} \qquad L_p(x)f^\Delta(x) \sim \frac{1}{\ell_p(x)\ell_{p+1}(x)}\frac{\ln|f|}{\ln|\ln|f||}.$$

Taking logarithms whenever appropriate, we then get:

1. if $\gamma_2(f) > \gamma_2(\ell_{p-1}(x))$ then $\gamma_1\left(L_p f^\Delta\right) = \gamma_1(\ln f)$;

2. if $\gamma_2(f) < \gamma_2(\ell_{p-1}(x))$ then $\ln\left(L_p f^{\Delta}\right) \sim -\ell_{p+1}(x)$;
3. if $\gamma_2(f) = \gamma_2(\ell_{p-1}(x))$ and $\ell_2(f) \not\sim \ell_{p+1}(x)$ then $\gamma_1\left(L_p f^{\Delta}\right) = \gamma_1(\ell_p(x))$;
4. if $\ell_2(f) \sim \ell_{p+1}(x)$ then $\gamma_1\left(L_p f^{\Delta}\right) < \gamma_1(\ell_p(x)) = \gamma_1(\ln f)$.

## 4. Nested forms

For an integer $m \geq 0$, reals $d \geq 0$ and $\epsilon > 0$ and a function $\phi$ such that $\gamma_1(\phi) < \gamma_1(\ell_m(x))$, consider $f = \ell_m^d(x)\phi(x)$ and $g = \ell_m^{d+\epsilon}(x)$. Then $\ln(f/g) = \ln\phi(x) - \epsilon\ell_{m+1}(x) = -\epsilon\ell_{m+1}(x) + o(\ell_{m+1}(x))$ is negative infinite at $+\infty$. Thus $f = o(g)$. If similarly $\epsilon < 0$, $g = o(f)$. In view of the previous relations, we define a *partial nested form* for a function $f$ that is infinite at $+\infty$ as an expression of the form

$$f = e_s\left(\ell_m^d(x)\phi\right) \qquad \text{where } s, m \geq 0,\ d \in \mathbb{R}^+ \text{ and } \gamma_1(\phi) < \gamma_1(\ell_m(x)).$$

Not every function $f$ in a Hardy field admits a nested form. In fact, if $f \to +\infty$, then either $\gamma_2(f) > \gamma_2(e_s(x))$ for all $s \geq 0$; or for all $m \geq 0$, there exists $f_m \in \mathbb{R}\langle\langle f\rangle\rangle$ such that $f_m \sim \ell_m(x)$; or else $f$ admits a partial nested form $f = e_s\left(\ell_m^d(x)\phi\right)$. The first two cases are strange situations in which, for example, $f$ cannot satisfy an algebraic differential equation over $\mathbb{R}$. In the third case, $\mathbb{R}\langle\langle f\rangle\rangle$ contains an element asymptotic to $\phi$. So the previous result applies to one of $\pm\phi^{\pm}$ and the function $\phi$ may in turn admit a partial nested form. Continuing in this way, we produce a sequence of $\phi_i$'s which either stops on a $\phi_i$ for which the second case above holds, or is an infinite sequence with decreasing $\gamma_1$, or contains a $\phi_i$ asymptotic to a non-zero constant $A$.

This gives a recursive definition of a *nested form*: a function $f$ has a nested form if there exists a finite sequence of $\phi_i$'s, $i = 0, \ldots, n$, with $\phi_0 = f$, such that each $\phi_{i+1}$ is the function $\phi$ which appears in a partial nested form of $\phi_i$ and $\phi_n = A + o(1)$ for a non-zero real $A$. (A few other technical constraints are added to ensure the uniqueness of the nested form of a function, in the case of existence.) For example, the following are two nested forms:

$$e_1\left(\ell_2^2(x)e_2\left(\ell_5^{1/3}(x)(2+o(1))\right)\right), \qquad \text{and} \qquad -e_1^{-1}\left(x^{\pi}\ell_1(x)e_2\left(\ell_5^{\sqrt{2}}(x)(13+o(1))\right)\right).$$

As a consequence to the previous results, if $f$ belongs to a Hardy field and $\Upsilon_1(\mathbb{R}\langle\langle f\rangle\rangle)$ is well ordered, then $f$ has a nested form. In particular, when $f$ satisfies an algebraic differential equation over $\mathbb{R}$ and belongs to a Hardy field, then $\Upsilon_1(\mathbb{R}\langle\langle f\rangle\rangle)$ is finite, with cardinality bounded by the order of the differential equation. It follows that only finitely many nested forms are possible for such solutions, and that those possible forms can be listed.

## 5. Further expansions

Nested forms only give a certain amount of asymptotic information. In particular, nothing is known about the $o(1)$. In certain cases, a possibility is to give a nested form representation for this $o(1)$, and continue recursively. We then get *nested expansions*, which extend asymptotic expansions, with no presumption of convergence. Nested expansions are guaranteed to exist for solutions of algebraic differential equations in a Hardy field and yield a unique representation. In particular, nested expansions can be calculated for exp-ln functions (modulo an oracle for constants); and for Liouvillian functions, although there is difficulty specifying *which* integral or algebraic function is under consideration. There are also known algorithms to add and multiply nested expansions, but this may be awkward. On the other hand the functional inverse of a nested expansion can be easily computed by an algorithm due to Salvy and Shackell.

## PART II: PROSPECTIVE RESULTS

### 6. Star products

In view of the identities $ab = \exp(\ln a + \ln b)$ and $ab = \ln\left(e^a e^b\right)$, we define the *star products* $*_k$ by

$$a *_k b = e_k(\ell_k(a) + \ell_k(b)) = e_{k-1}(\ell_{k-1}(a)\ell_{k-1}(b)), \qquad \text{for } k \in \mathbb{Z}.$$

We have $a *_0 b = a + b$, $a *_1 b = ab$. For $r \in \mathbb{R}$, we also define $a^{*_k^r} = e_k(r\ell_k(a))$. These definitions yield properties which give star products their name: each $*_k$ is commutative and associative; $*_{k+1}$ is left and right distributive over $*_k$; $*_{k+1}$ admits $e_{k+1}(0)$ as a neutral element and $e_k(0)$ as a zero.

Sample expressions that involve star products are:

$$x *_2 \ell_1(x) = \exp(\ell_1(x)\ell_2(x)), \qquad (\exp x)^{*_2^2} = \exp\left(x^2\right), \qquad e^x *_{-1} x = \ln\left(e_2(x) + e^x\right),$$

$$x^{*_{-1}^2} = \ln\left(2e^x\right) = x + \ln 2, \qquad (\exp x)^{*_3^2} *_2 \ell_1(x) = e_2\left(\ell_1^3(x)\right) *_2 \ell_1(x) = e_1\left(e_1\left(\ell_1^3(x)\right)\ell(2(x)\right).$$

Of course, any exp-ln function can be written as an expression in the real constants and the functions $e_n$ for $n \in \mathbb{Z}$ using star products $*_k$ for $k \in \mathbb{Z}$ only.

The advantage of star products is their nice behaviour with the $\gamma_k$'s: if $a$ and $b$ have infinite limit at $+\infty$, then $\gamma_{k-1}(a*_k b) > \max\{\gamma_{k-1}(a), \gamma_{k-1}(b)\}$ and $\gamma_k(a*_k b) = \max\{\gamma_k(a), \gamma_k(b)\}$. Furthermore, taking $*_k$ powers does not affect asymptotic relations between $e_s(x)$'s. In view of these results, we expect to find a better presentation of nested expansions in terms of star products and hope for simpler algorithms to deal with nested expansions.

### 7. Partial Hardy fields

We say that a ring $\mathcal{H}$ of functions of two variables $x$ and $y$ is a *partial Hardy field* if

1. for sufficiently large $y_0$, $\{f(x, y_0) \mid f \in \mathcal{H}\}$ is a Hardy field;
2. for sufficiently large $x_0$, $\{f(x_0, y) \mid f \in \mathcal{H}\}$ is a Hardy field;
3. for $f \in \mathcal{H}$, $\lim_{y\to+\infty} f(x, y)$ is either identically $\pm\infty$ or else an element of a Hardy field;
4. for $f \in \mathcal{H}$, $\lim_{x\to+\infty} f(x, y)$ is either identically $\pm\infty$ or else an element of a Hardy field.

The point of this definition is to allow for *multivariate nested expansions.*

Let $f(x, y)$ be a function in a partial Hardy field. Provided $\mathbb{R}\langle\langle f\rangle\rangle$ is a partial Hardy field, the sequence of $\phi_i(x)$ obtained by taking the nested expansion of $f(x, y_0)$ for fixed $y = y_0$ becomes independent of $y_0$ for sufficiently large $y_0$. Thus with an assumption of well-orderedness, we obtain a first nested expansion,

$$f(x, y) \underset{x\to+\infty}{\sim} e_{s_0}\left(\ell_{m_0}^{d_0}(x)e_{s_1}^{\pm 1}\left(\cdots e_{s_n}^{\pm 1}\left(\ell_{m_n}^{d_n}(x)(\phi(y) + o(1))\right)\cdots\right)\right),$$

where by point (4) of the definition, $\phi$ belongs to a Hardy field. Now, with reasonable assumptions, $\phi$ admits a nested form.

Once again, if $f$ satisfies an algebraic partial differential equation, there is a limitation on the nested forms allowed to occur. However, a complication is that solutions of PDE's contain arbitrary functions, which can be set by specifying the limiting behaviour of $f$ in one direction.

### Bibliography

[1] Bourbaki (N.). – *Éléments de Mathématiques*, Chapter V: Fonctions d'une variable réelle (appendice), pp. 36–55. – Hermann, Paris, 1961, second edition.

# Asymptotics of Implicit Functions and Computer Algebra

*Bruno Salvy*
Inria Rocquencourt

September 9, 1996

[summary by Joris Van der Hoeven]

**Abstract**

We describe several algorithms for the asymptotic inversion of functions, from the computer algebra viewpoint. We start with some classical results, such as the inversion theorem of Lagrange. We next consider functions which can only be expanded in more complicated asymptotic scales. Finally, we briefly discuss the problem of finding the asymptotic behaviour of implicit functions.

## 1. Introduction

Let $f$ be a sufficiently regular function which admits a functional inverse $f^{inv}$ in the neighbourhood of some point. One often encounters the problem of determining the asymptotic behaviour of $f^{inv}$ in terms of the asymptotics of $f$. For instance, this problem arises systematically when computing integral transforms by the saddle-point method. If the function admits a power series expansion near the point of interest, then Lagrange's inversion theorem can be used, or Brent and Kung's fast algorithm for power series inversion (see section 2).

However, one often has to deal with logarithmic or exponential singularities, whence a more general device for asymptotic functional inversion is needed. A first approach would be to consider more general classes of singularities of a particular form. For instance, one may consider functions which admit an asymptotic expansion of the form

$$f = e^x x^{-\alpha} \sum_{n \geq 0} d_n x^{-n} \qquad (x \to \infty);$$

see section 3. More generally, one might consider the class of exp-log functions, i.e. functions built up from $\mathbb{Q}$ and $x$ by $+, -, \times, /, \exp$ and $\log$. A first algorithm to obtain asymptotic information about functional inverses of such functions—in the form of a so-called *nested expansion*—was given by Salvy and Shackell [6]; see also section 4.

An interesting more general problem than the inversion problem is to find an algorithm to determine the asymptotic behaviour of implicit functions, say the solutions of exp-log functions in two variables. A partial algorithm to determine the potential nested expansions of such functions was proposed in [7] and will be discussed in section 5.

More generally, one might require complete asymptotic expansions of the inverses of exp-log functions and solutions of implicit equations. A nice algorithm to compute complete asymptotic expansions of exp-log functions themselves was given in [4]. Notice also that extending the techniques from [4], solutions for the inversion and implicit function problems for exp-log functions were given independently in [10] and [11].

## 2. Classical results

Let $f(y) = f_0 + f_1 y + f_2 y^2 + \cdots$ be a power series with $f_0 \neq 0$ and assume that $y(x)$ satisfies

$$y = x f(y).$$

Then Lagrange's inversion theorem gives a formula for $[x^n]y$:

$$[x^n]y = \frac{1}{n}[y^{n-1}]f^n.$$

The formula can be generalized to

$$[x^n]G = \frac{1}{n}[y^{n-1}]G'f^n,$$

for an arbitrary power series $G$. If $f$ resp. $G$ have an analytic meaning (i.e. the power series converge, or the power series are the asymptotic expansion of some functions), then the same holds for the above formulae.

EXAMPLE. The generating function $y$ of Cayley trees satisfies

$$y = xe^y.$$

Hence $[x^n]y = \frac{1}{n}[y^{n-1}]e^{ny} = \frac{n^{n-1}}{n!}$. Other trees can be treated in a similar way.

Assume now that we do not merely want the $n$-th coefficient of the inverse of a power series $x(y) = y + a_2 y^2 + a_3 y^3 + \cdots = F(y)$, but the first $N$ coefficients of $y(x)$. This problem can be reduced to the problem of functional composition by the Newton method: define the sequence $y_n(x)$ by $y_0 = x$ and

$$y_{n+1} = y_n - \frac{F(y_n) - x}{F'(y_n)}.$$

Brent and Kung have shown that the number of correct terms doubles at each step [1]. Suitably truncating the Newton iterations then yields an inversion algorithm for power series of the same complexity as functional composition, i.e. $O(M(N)\sqrt{N \log N}) = O((N \log N)^{3/2})$ [1].

EXAMPLE. How to find the $n$-th positive root $x_n$ of $\tan x = x$? Looking at the graph, we see that $x_n = (2n+1)\frac{\pi}{2} - u_n$ with $u_n \to 0$. Applying tan we get

$$(2n+1)\frac{\pi}{2} - u_n = \tan\left(\frac{\pi}{2} - u_n\right),$$

whence

$$\frac{1}{u_n + \tan(\frac{\pi}{2} - u_n)} = \frac{2}{(2n+1)\pi} = t_n,$$

with $t_n \to 0$. Inverting the above power series, we get

$$x_n = \pi n + \frac{\pi}{2} - \frac{1}{\pi n} + \frac{1}{2\pi n^2} - \left(\frac{1}{4\pi} + \frac{2}{3\pi^3}\right)\frac{1}{n^3} + \cdots.$$

## 3. Inverses of more general functions

In this section we are interested in computing the asymptotic inverse of a function which admits

$$e^y y^{-\alpha} D(y^{-1}) \qquad (d_n \neq 0, y \to \infty)$$

as an asymptotic expansion, where $D(y^{-1}) = \sum_{n\geq 0} d_n y^{-n}$. In other words, we want the asymptotic solution to

$$e^y y^{-\alpha} D(y^{-1}) = x$$

in $y$. A first method [2] is to do this is to take logarithms

$$y = \log x + \alpha \log y - \log D(y^{-1}), \tag{1}$$

and to replace $y$ by the right hand side in an iterative manner. This yields an expansion of the form

$$y \approx \log x + \sum_{n\geq 0} \frac{P_n(\log\log x)}{\log^n x},$$

where the $P_n$ are polynomials.

A faster method to compute the $P_n$ was proposed in [5]. Setting $\zeta = \log\log x$ and $t = (\log x)^{-1}$, we set

$$y - \log x = P(\zeta, t) = \sum_{n\geq 0} P_n(\zeta) t^n.$$

Taking logarithms, this yields

$$\log y = \log(\log x + P) = \zeta + \log(1 + tP).$$

From (1), we get

$$P = \alpha\zeta + \alpha \log(1 + tP) - \log D\left(\frac{t}{1+tP}\right).$$

In this equation $\zeta$ and $t$ are decoupled. We get

$$P(0,t) = \alpha \log(1 + tP(0,t)) - \log D\left(\frac{t}{1+tP(0,t)}\right)$$

and

$$\left(\frac{1}{\alpha} - t\right)\frac{\partial P}{\partial \zeta} + t^2 \frac{\partial P}{\partial t} = 1.$$

These two relations enable us to compute the coefficients of $P$ efficiently.

EXAMPLE. The number of prime numbers $\pi(x)$ smaller than $x$ satisfies

$$\pi(x) \approx \frac{x}{\log x}\left(1 + \frac{1!}{\log x} + \frac{2!}{\log^2 x} + \cdots\right).$$

Using the above method we find the asymptotic expansion

$$p(n) \approx n \log n \left(1 + \frac{\log\log n - 1}{\log n} + \frac{\log\log n - 2}{\log^2 n} + \cdots\right)$$

for the $n$-th prime number $p(n)$.

The above result can be extended to get asymptotic expansions for $\log y$ and $e^{\beta y} y^{\gamma} G(y^{-1})$ [5].

## 4. Inversion of exp-log functions

Let $f$ be an exp-log function (i.e. a function built up from $\mathbb{Q}$ and $x$ by $+, -, \times, /$, exp and log), which tends to infinity for $x \to \infty$. We denote by $\exp_i$ resp. $\log_i$ the $i$-th iterate of exp resp. log. We write $f \prec\!\!\prec g$, if $f$ is of a smaller comparability class than $g$, i.e. $(\log|f|)/(\log|g|)$ tends to 0.

In this section, we present an algorithm to compute a "nested expansion" for the asymptotic functional inverse of $f$. Such a nested expansion often (although not always) yields an asymptotic expansion of the function. In any case, the limit behaviour of a function can be determined from its nested expansion.

One first defines a *nested form* as being a finite sequence $\{(s_i, \varepsilon_i, m_i, d_i, \phi_i)\}$ for $1 \le i \le n$. Here $s_i, m_i \in \mathbb{N}, d_i \in \mathbb{R}, \varepsilon_i = \pm 1$, $\phi_i$ lies in a Hardy field and

$$\phi_{i-1}(x) = \exp_{s_i}^{\varepsilon_i}(\log_{m_i}^{d_i}(x)\phi_i(x)),$$

with $\phi_i \prec\!\!\prec \log_{m_i}$ for $2 \le i \le n$. We also require that

- $\phi_n$ has a finite limit $l$;
- $d_n \neq 1$ unless $s_n = 0$ or $m_n = 0$;
- $d_i > 0$ unless $s_i = 0$.

Continuing the process of taking nested forms with $\phi_n - l$ (if possible), we obtain a *nested expansion*.

EXAMPLE. $\exp[\log^2 x \exp[\sqrt{\log\log x}(7 + \phi(x))]]$ is a nested form, if $\phi(x)$ tends to 0.

John Shackell was the first to prove in [8] that any exp-log function admits a nested form, and even a nested expansion. Nested forms and expansions are particularly well suited to the purpose of functional inversion. The algorithm proceeds as follows:

*Algorithm.* An exp-log function $f$ tending to infinity is given at input, and we compute a nested expansion of its functional inverse at infinity.

1. Rewrite $f$ as a NF
$$f(y) = \exp_s(\log_m^d(y) f_1(y)) = x.$$
2. Invert
$$y = \exp_m(\log_s^{1/D} x g_1(x)). \qquad (E)$$
3. Iterate
   - Compute $NF(g_i) = \exp_{s_i}^{\varepsilon_i}(\log_{m_i}^{d_i} yG_i(y))$,
   - Substitute $(E)$ in $\log_{m_i}^{d_i} y$,
   - Rewrite to get $g_{i+1}$,

   until $NF(g_i) = c + \varepsilon(y)$, with $\varepsilon(y) \to 0$, yielding $NF(y(x))$.
4. Repeat the above procedure for $\varepsilon(y)$, whence $NE(y(x))$.

EXAMPLE.

$$f(y) = ye^{\log^2 ye^{\sqrt{\log\log y}}} = x.$$

Step 1: rewrite $f$ as NF

$$\exp\left[\log^2 y \exp[\sqrt{\log_2 y}(1 + W)]\right],$$

with

$$W = \log_2^{-1/2} y \log(1 + \log^{-1} ye^{-\sqrt{\log_2 y}}).$$

Step 2: Invert

$$y = \exp\left[\sqrt{\log x}\exp\left(-\frac{1}{2}\sqrt{\log_2 y}(1+W)\right)\right].$$

Step 3: Iterate

$$\sqrt{\log_2 y} = \frac{1}{\sqrt{2}}\sqrt{\log_2 x}\left(1+\frac{1+W}{2\sqrt{\log_2 y}}\right)^{-1/2},$$

whence the nested form for $y$:

$$y = \exp\left[\sqrt{\log x}\exp\left(-\frac{1}{2\sqrt{2}}\sqrt{\log_2 x}(1+\varepsilon(x))\right)\right].$$

Step 4: Repeat

$$y = \exp\left[\frac{\sqrt{\log x}}{e^{\frac{1}{2\sqrt{2}}\log_2 x}}\cdot e^{1/8}\cdot\left(1-\frac{1}{32\sqrt{2}}\log_2^{-1/2} x+\right.\right.$$
$$\left.\left.\frac{1}{4096}\log_2^{-1} x+\frac{383}{393216\sqrt{2}}\log_2^{-3/2} x+\cdots\right)\right].$$

## 5. Asymptotic expansion of implicit functions

Assume now that $H$ is an exp-log function in two variables $x$ and $y$. By theorems of Khovanskii and van den Dries [3, 9], if $y(x)$ is a real solution of $H(x,y)=0$ (for $x\to\infty$), then (the germ of) $y(x)$ belongs to a Hardy field. In particular, $y(x)$ does not present any oscillatory phenomena at infinity. As a corollary [7], $y(x)$ has a nested form and even a nested expansion. A question is how to compute all possible nested expansions of such solutions.

The idea from the algorithm in [7] is to compute the asymptotic behaviour for $H(x,y)$ for all possible asymptotic behaviours of $x$ and $y$. In order to do so, one often has to distinguish several number of cases, but always finitely many, depending on the values of $x$ and $y$. The strategy is best illustrated on an example.

EXAMPLE.

$$H(x,y) = \frac{\exp(x^2+2x\log^2 x+y)}{\exp(x^2+x\log^2 x+y)} - 1.$$

Three cases are distinguished:

$$H \sim \begin{cases} 1, & \text{if } y\to-\infty \text{ and } x \ll y;\\ ?, & \text{if } \log|y| \sim 2\log x;\\ \exp(x\log^2 x), & \text{otherwise.}\end{cases}$$

Since $H=0$, we find that $\log|y|\sim 2\log x$, whence $y\asymp -x^2$. Continuing the process, we find

$$y = -x^2 - 2x\log^2 x + \frac{e^{-x\log^2 x}}{x^2} - \frac{e^{-2x\log^2 x}}{2x^2} + \cdots.$$

We remark that it is not claimed that the obtained nested expansions are indeed nested expansions of actual solutions. For a solution to this problem, we refer to [11].

## Bibliography


[1] Brent (R. P.) and Kung (H. T.). – Fast algorithms for manipulating formal power series. *Journal of the ACM*, vol. 25, 1978, pp. 581–595.

[2] De Bruijn (N. G.). – *Asymptotic Methods in Analysis.* – Dover, 1981. A reprint of the third North Holland edition, 1970 (first edition, 1958).

[3] Khovanskii (A. G.). – Fewnomials and Pfaff manifolds. In *Proceedings of the International Congress of Mathematicians*, pp. 549–564. – 1983.

[4] Richardson (Dan), Salvy (Bruno), Shackell (John), and Van der Hoeven (Joris). – Asymptotic expansions of exp-log functions. In Lakshman (Y. N.) (editor), *Symbolic and Algebraic Computation.* pp. 309–313. – New York, 1996. Proceedings ISSAC'96. Zürich.

[5] Salvy (Bruno). – Fast computation of some asymptotic functional inverses. *Journal of Symbolic Computation*, vol. 17, 1994, pp. 227–236.

[6] Salvy (Bruno) and Shackell (John). – Asymptotic expansions of functional inverses. In Wang (Paul S.) (editor), *Symbolic and Algebraic Computation.* pp. 130–137. – New York, 1992. Proceedings of ISSAC'92, Berkeley.

[7] Salvy (Bruno) and Shackell (John). – *Symbolic Asymptotics: Functions of Two Variables, Implicit Functions.* – Research Report n° 2883, Institut National de Recherche en Informatique et en Automatique, 1996. Submitted to the *Journal of Symbolic Computation.*

[8] Shackell (John). – Growth estimates for exp-log functions. *Journal of Symbolic Computation*, vol. 10, n° 6, December 1990, pp. 611–632.

[9] Van den Dries (Lou). – Analytic Hardy fields and exponential curves in the real plane. *American Journal of Mathematics*, vol. 106, 1984, pp. 149–167.

[10] van der Hoeven (J.). – *General algorithms in asymptotics II: Common operations.* – Technical Report n° LIX/RR/94/10, LIX, École polytechnique, France, 1994.

[11] Van der Hoeven (Joris). – *Asymptotique Automatique.* – PhD thesis, École polytechnique, Palaiseau, France, 1997.

# Part 3

# Analysis of Algorithms and Data Structures

# Counting Polynomials over Finite Fields and Analysis of Algorithms

*Daniel Panario*
University of Toronto

October 7, 1996

[summary by Mireille Régnier]

## 1. Motivation

In this talk, we comment on several problems in finite fields, and their relation with analytic combinatorics. Algebraic algorithms that deal with polynomials over finite fields can often be analyzed by counting polynomials with particular properties. We show that the most important characteristics of these algorithms can be treated systematically by a methodology based on generating functions and asymptotic analysis. We focus on three problems: polynomial factorization, irreducibility tests for polynomials, and discrete logarithm. For each problem, we present an efficient algorithm, we derive interesting counting expressions, and we mention known results.

## 2. Basic methodology

**2.1. Generating functions.** Let $\Phi$ be a class of monic polynomials, $\chi$ some integer-valued parameter on $\Phi$. Let

$$\Phi(z,u) = \sum_{\omega \in \Phi} z^{|\omega|} u^{\chi(\omega)}.$$

The coefficient $[z^n u^k]\Phi(z,u)$ represents the number of polynomials of degree $n$ with $\chi$-parameter equal to $k$. Averages and standard deviations are obtained by taking successive derivatives of bivariate generating functions with respect to $u$, then setting $u = 1$. For instance, the mean is:

$$\frac{[z^n] \left.\frac{\partial P(z,u)}{\partial u}\right|_{u=1}}{[z^n]P(z,1)} = \frac{p_n'(1)}{p_n(1)}.$$

**2.2. Asymptotic analysis.** Generating functions encode exact informations on their coefficients. Their behavior near their dominant singularity is an important source of coefficient asymptotics.

A first method is known as singularity analysis due to Flajolet & Odlyzko. This requires analytic continuation (isolated singularity). However, there are some problems in which the generating functions present a natural boundary at $|z| = 1$ (each point at the unit circle is singular). Darboux's method could be used as an alternative in these cases. Finally, in some cases we use also a saddle point approximation.

**2.3. Permutation model.** The joint distribution of degrees in the prime decomposition of a random polynomial over $\mathbb{F}_q$ having degree $n$ admits as a limit, when $q \to \infty$ ($n$ staying fixed!), the joint distribution of cycle lengths in random permutations of size $n$. This gives rise to a useful heuristic: *probabilistic properties of polynomial factorization often have a shape resembling that of corresponding properties of the cycle decomposition of permutations to which they usually reduce as* $q \to \infty$.

## 3. Factoring polynomials over finite fields

The results in this part of the talk are from [1]. The Polynomial factorization algorithm proceeds in three steps:

**ERF:** *Elimination of repeated factors* replaces a polynomial by square-free ones that contain all the irreducible factors of the original polynomial with exponents reduced to 1.

**DDF:** *Distinct-degree factorization* splits a square-free polynomial into a product of polynomials whose irreducible factors all have the same degree.

**EDF:** *Equal-degree factorization* factors a polynomial whose irreducible factors have the same degree.

As our interest is in *dominant asymptotics*, we restrict our attention to the costs of products and gcd's that we assume to have constant costs $\tau_1$ and $\tau_2$ respectively.

**3.1. Elimination of repeated factors (ERF).** The first step in the factorization chain of a polynomial is the *elimination of repeated factors* (ERF). One proves that:

THEOREM 1. *(i) A random polynomial of degree $n \geq 2$ in $\mathbb{F}_q[x]$ has a probability $1 - 1/q$ to be square-free.*

*(ii) The degree of the non-square-free part of a random polynomial has expected value asymptotic to*

$$C_q = \sum_{n \geq 1} \frac{n I_n}{q^{2n} - q^n},$$

*where $I_n$ is the number of irreducible polynomials of degree $n$, and a geometrically decaying probability tail. When $q \to \infty$, then $C_q \sim 1/q$.*

Consequently, the overall cost of the recursive calls in the elimination of repeated factors remains $O(1)$ on average; alternative strategies giving the full square-free factorization will lead to asymptotically equivalent costs; the ERF phase has a cost dominated by that of its first gcd.

THEOREM 2. *The expected cost of the ERF phase applied to a random polynomial of degree $n$ satisfies*

$$\overline{ERF}_n \sim \tau_2 n^2.$$

**3.2. Distinct-degree factorization (DDF).** The second stage of our factorization algorithm requires finding the *distinct-degree factorization* (DDF) of the square-free polynomial $a$, i.e., splitting $a$ under the form $b_1 \cdots b_n$ where $b_k$ is the product of irreducible factors of $a$ of degree $k$. The algorithm is $O(n^3)$. We provide a precise comparison of three strategies for the DDF phase: the basic rule, the "half-degree" rule and the "early abort" rule. The global saving of the early abort rule is of 36%, and the expected cost of $O(\log q \cdot n^3)$ for DDF clearly dominates the whole factorization chain.

**3.3. Equal-degree factorization (EDF).** DDF does not completely factor a polynomial that has different factors of same degree.

THEOREM 3. *(i) The probability that DDF yields the complete factorization is asymptotic to*

$$c_q = \prod_{n \geq 1} \left(1 + \frac{I_n}{q^n - 1}\right) (1 - q^{-n})^{I_n},$$

$$c_2 \doteq 0.6656,\ c_{257} \doteq 0.5618,\ c_\infty = e^{-\gamma} \doteq 0.5614.$$

*(ii) The degree of the part of the polynomial that remains to be factored by the EDF algorithm is asymptotic to* $\log n$.

The factorization problem is reduced to factoring polynomials $b_k$ that have all their irreducible factors of the same (known) degree $k$. Our reference is Cantor-Zassenhaus' probabilistic algorithm.

Each factor of $b$ has probability $\alpha = (q-1)/(2q)$ to be a factor of $d$, and probability $\beta = (q+1)/(2q)$ to divide $b/d$. A random choice splits $b$ in $\langle \ell, j-\ell \rangle$ factors with Bernoulli probability $\binom{j}{\ell}\alpha^\ell \beta^{j-\ell}$. The analysis combines a recursive partitioning problem akin to digital tries with estimates on the degree of irreducible factors of random polynomials.

THEOREM 4. *The expected cost of the EDF phase satisfies*

$$\overline{EDF}_n \sim \frac{\tau_1}{\alpha\beta} \sum_{k=1}^{[n/2]} \mu_k, \qquad \mu_k = \left\lfloor \log_2(q^k-1)/2 \right\rfloor + \nu(q^k-1)/2 - 1.$$

*In addition, this cost is $O(n^2)$, and for $-1/3 \le \xi_n \le 1/3$,*

$$\overline{EDF}_n \sim \left( \frac{3}{4}\tau_1 \frac{q^2}{q^2-1} \log_2 q \cdot n^2 \right) (1 + \xi_n + o(1)).$$

## 4. Irreducibility tests for polynomials

A fundamental problem in finite fields is the construction of extension fields, that may be done by using an irreducible polynomial over the ground field with degree equal to the degree of the extension. Therefore, finding irreducible polynomials is a central problem in finite fields. A probabilistic algorithm is presented in [5]. The central idea is to take polynomials at random and test them for irreducibility. This suggests the study of the probability that a random polynomial of degree $n$ contains no irreducible factors of degree up to certain value $m$ (such polynomials are called $m$-rough). Gao and Panario [2] considered the case $m = O(\log n)$ and proved:

THEOREM 5. *Denote by $P_q(n,m)$ the probability of a random monic polynomial of degree $n$ over $\mathbb{F}_q$ being $m$-rough. Then when $n \to \infty$ and uniformly for $q$ and $1 \le m \le O(\log n)$,*

$$P_q(n,m) = \prod_{k=1}^{m} \left(1 - \frac{1}{q^k}\right)^{I_k} (1 + o(1)),$$

THEOREM 6. *Let $g_q(m) = \prod_{k=1}^{m} \left(1 - \frac{1}{q^k}\right)^{I_k}$. Then, for any prime power $q$ and positive integer $m$,*

$$e^{-H_m} \le g_q(m) \le \left(1 - \frac{1}{\sqrt{q}}\right)^{-\frac{q}{q-1}} e^{-H_m}.$$

*When $q \to \infty$, we have*

$$g_q(m) = \prod_{k=1}^{m} \left(1 - \frac{1}{q^k}\right)^{I_k} \to e^{-H_m} \sim \frac{e^{-\gamma}}{m},$$

*where $\gamma$ is Euler's constant and $e^{-\gamma} = 0.56416\ldots$.*

## 5. Discrete logarithm problem

For any element $b \in \mathbb{F}_q$, $b \neq 0$, there exists an integer $x$, $0 \le x \le q-2$, such that $b = \alpha^x$, where $\alpha$ is a generator. We call $x$ the *discrete logarithm* of b in the base $\alpha$.

We present here the *index calculus algorithm* to compute the discrete logarithm of any $b \in \mathbb{F}_q$, $b \neq 0$ and restrict ourselves to the case of $\mathbb{F}_{2^n}$.

This method consists of two parts. First, one builds a large database of logarithms by finding the logarithms of all irreducible polynomials of degree at most $m$, where $m$ is a fixed positive integer. Second, one computes individual logarithms. To compute the logarithm of an element $g \in \mathbb{F}_{2^n}$, $g \neq 0$, one takes a random integer $a$, computes $h = g \cdot \alpha^a$, where $\alpha$ generates $\mathbb{F}_{2^n}$ and factors $h$ in $h = \prod_{i=1}^{t} p_i^{e_i}$. If each irreducible factor $p_i$ has degree $p_i \leq m$, then

$$\log g = \sum_{i=1}^{t} e_i \log p_i - a,$$

which can be easily evaluated by looking up in the database. If not all $p_i$ have degree $\leq m$, then one generates another integer $a$ and repeats.

THEOREM 7 ([4]). *Let $\mathbb{F}_q$ be fixed, $f_m(z) = \prod_{k=1}^{m}(1 - z^k)^{-I_k}$, and $r_0 = r_0(n, m)$ be the unique solution in $(0, 1)$ of the equation $r_0(f_m'/f_m)(r_0) = n$, and let*

$$b(r_0) = \left(\frac{f_m{}'}{f_m}(r)\right)'\Bigg|_{r=r_0}.$$

*Then, for*

$$(\log n)(\log\log n)^{-1} \leq m \leq n \log\log n (\log n)^{-1}, \qquad n \to \infty,$$

$$[z^n] f_m(z) = (1 + o(1)) \frac{f_m(r_0) r_0^{-n}}{\sqrt{2\pi b(r_0)}}.$$

Soundararajan (1995) completed the full range of $m$ estimating $[z^n]f_m(z)$ using recurrences relations. This could be done using partial fraction expansions for $1 \leq m \leq (\log n)(\log\log n)^{-1}$, and singularity analysis for $n \log\log n (\log n)^{-1}$.

## 6. Conclusions

Generating functions and singularity analysis allow for counting random polynomials over finite fields. We applied this methodology to give precise average-case analysis of a complete polynomial factorization algorithm [1]. Using this methodology, von zur Gathen, Gourdon & Panario (work in progress) present further research related to the average-case analysis of polynomial factorization algorithms. This work centers around [3], and the factoring algorithms of the 90's.

We also commented on other problems using polynomials over finite fields: tests and constructions of irreducible polynomials [2]; discrete log in $\mathbb{F}_{2^n}$ [4]; (see also Panario & Viola, work in progress).

## Bibliography


[1] Flajolet (Philippe), Gourdon (Xavier), and Panario (Daniel). – *Random Polynomials and Polynomial Factorization.* – Research Report n° 2852, Institut National de Recherche en Informatique et en Automatique, March 1996. To appear in the *Proceedings of ICALP'96, Lecture Notes in Computer Science*.

[2] Gao (Shuhong) and Panario (Daniel). – Density of normal elements. *Finite Fields and their Applications*, vol. 3, n° 2, 1997, pp. 141–150.

[3] Kaltofen (E.) and Shoup (V.). – Subquadratic-time factoring of polynomials over finite fields. *Mathematics of Computation*, April 1998. – To appear.

[4] Odlyzko (A. M.). – Discrete logarithms in finite fields and their cryptographic significance. In *Advances in cryptology. Lecture Notes in Computer Science*, vol. 209, pp. 224–314. – Berlin, 1985. Proceedings of a conference held in Paris, 1984.

[5] Rabin (Michael O.). – Probabilistic algorithms in finite fields. *SIAM Journal on Computing*, vol. 9, n° 2, 1980, pp. 273–280.

# Distribution of Image Points in Random Mappings

*Michèle Soria*
Université Paris VI

November 10, 1996

[summary by Pierre Nicodème]

**Abstract**

This talk presents a general theorem which can be used to identify the limiting distribution for a class of combinatorial schemata. For example, many parameters in random mappings can be covered in this way.

## 1. Methods

We consider the general working scheme "Symbolic Structures $\mathcal{A}$ or $\{\mathcal{A}, \omega\} \to$ Generating Functions $a(z)$ or $a(u,z) \to a_n$ or $a_{n,k}$". Then by Cauchy's formula, we get for structures $\mathcal{A}$

$$a(z) = \sum_{\alpha \in \mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n \geq 0} a_n \frac{z^n}{n!} \Longrightarrow \frac{a_n}{n!} = \frac{1}{2i\pi} \oint a(z) \frac{dz}{z^{n+1}}.$$

When considering marked structures with parameters $\{\mathcal{A}, \omega\}$, ($\omega$ is a mapping $\mathcal{A} \to \mathbb{N}$), we have

$$a(u,z) = \sum_{\alpha \in \mathcal{A}} u^{\omega(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n,k} a_{n,k} u^k \frac{z^n}{n!}.$$

In this case, $a_{n,k}$ can be obtained by double Cauchy inversion, or by Cauchy inversion and Continuity Theorem. Table 1 gives some examples of translation of marked combinatorial structures to generating functions. The mark is represented by character "$\bullet$" and translated to parameter $u$.

By a classical theorem about characteristic functions $(X_n)$ converges weakly to $Y$ if and only if $\phi_{X_n}(\theta)$ converges to $\phi_Y(\theta)$ for all $\theta$, with $\phi_Z = E(e^{i\theta Z})$. We also have $a(u,z) = \sum_{n,k} u^k \frac{z^n}{n!} = \sum_n p_n(u) \frac{z^n}{n!}$, which gives the probability generating function of $X_n$ as $p_n(u)/p_n(1) = \sum_n \Pr(X_n = k) u^k$. We refer to [2] for the concept of (labelled) combinatorial structures and their translation to generating functions.

| Description | Structure | Generating Function |
|---|---|---|
| Degree at the root in Cayley trees | $\mathcal{A} = \text{Node} \times \text{Set}(\bullet\mathcal{A})$ | $a(u,z) = z \exp(ua(z))$ |
| Random Mappings | $\mathcal{G} = \text{Set}(\text{Cycle}(\mathcal{A}))$ | $g(z) = \frac{1}{1-a(z)}$ |
| — by number of cycles | $\mathcal{G} = \text{Set}(\bullet\, \text{Cycle}(\mathcal{A}))$ | $g(u,z) = \exp\left(u \log \frac{1}{1-a(z)}\right)$ |
| — by number of trees | $\mathcal{G} = \text{Set}(\text{Cycle}(\bullet\mathcal{A}))$ | $g(u,z) = \frac{1}{1-ua(z)}$ |

TABLE 1. Some examples of generating functions

## 2. Trees and Random Mappings

A random mapping is an arbitrary mapping $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that every mapping has probability $n^{-n}$. A mapping $\phi$ can be identified to its functional graph $G_\phi$ with vertices $\{1, \ldots, n\}$ and edges $(i, \phi(i))$, for $1 \geq i \geq n$. Each component of $G_\phi$ consists of a cycle and every cyclic point is the root of a tree.

The basic property for analysis is that solutions of functional equations usually have algebraic singularity of square-root type. For trees, we get $a(u, z) = t(u, z) - h(u, z)\sqrt{1 - z/\rho(u)}$. For sequences of trees, we get an expression of the form $1/(1 - a(u, z))$, and for random mappings an expression of the form

$$s(u, z) = \frac{1}{1 - T(u, z) + h(u, z)\sqrt{1 - z/\rho(u)}}.$$

We recall that when we get an expression of the form $1/(1 - uC(z))$, the asymptotic distribution of the corresponding random variable depends on the value $C(\rho_c)$, where $\rho_c$ is the only singularity on the circle of convergence of $C(z)$. If $C(\rho_c) > 1$, the limit law is normal; if $C(\rho_c) < 1$, the limit law is derivative of geometric, and if $C(\rho_c) = 1$ the limit law is Rayleigh.

## 3. Examples

*Leaves.* For Cayley trees, we have $a(u, z) = ze^{a(u,z)} + z(u - 1)$, for sequences of trees, $s(u, z) = 1/(1 - a(u, z))$, and for functional graphs

$$g(u, z) = \frac{1}{1 - ze^{a(u,z)}} = \frac{1}{1 - a(u, z) + z(u - 1)}.$$

*Nodes of arity $r$.* For trees,

$$a(u, z) = z\left(\sum_{m \neq r} \frac{a^m(u, z)}{m!} + u\frac{a^r(u, z)}{r!}\right) = ze^{a(u,z)} + z(u - 1)\frac{a^r(u, z)}{r!}.$$

For sequences of trees, we have $s(u, z) = 1/(1 - a(u, z))$, and for functional graphs,

$$g(u, z) = \frac{1}{1 - a(u, z) + z(u - 1)\left(\frac{a^{r-1}}{(r-1)!} - \frac{a^r}{r!}\right)}.$$

*Nodes at distance $d$ from a cycle.* We have the recurrence

$$a_0(u, z) = ua(z), \qquad a_{d+1}(u, z) = ze^{a_d(u,z)}.$$

For functional graphs, this gives $g(u, z) = 1/(1 - a_d(u, z))$.

*Nodes with $r$ pre-images in total.* For trees, we have $a(u, z) = ze^{a(u,z)} + (u - 1)\alpha_{r+1}z^{r+1}$, where $\alpha_{r+1} = (r+1)^r$ is the number of trees of size $r+1$. For functional graphs, we have $\mathcal{G} = \mathrm{Set}(\mathrm{Cycle}(\mathcal{A}))$, which translates to $g(z) = \exp\left(\sum_{p \geq 0} \frac{a^p(z)}{p}\right)$. This gives

$$g(u, z) = \frac{1}{1 - ze^{a(u,z)}} \exp\left(\frac{z^r}{r}\sum_p (u^p - 1)\frac{r^{r-p}}{(r - p)!}\right) = \frac{K(u, z)}{1 - a(u, z) + (u - 1)\alpha_{r+1}z^{r+1}}.$$

*Nodes d iterated.* (These nodes are at distance $\geq d$ from a leaf.) For trees, we have

$$a_d(u,z) = xue^{a_d(u,z)} - (u-1)l_d(z) \qquad \text{with} \quad l_0(z) = 0, \quad l_{d+1}(z) = ze^{l_d(z)}.$$

For functional graphs, we have, for nodes at distance $\geq d$ of a leaf of their sub-tree, $s_d(u,z) = 1/(1-a_d(u,z))$. For nodes at distance $\geq d$ of a leaf, we have

$$g_d(u,z) = \frac{1}{1-uze^{a_d(u,z)}} = \frac{1}{1-a_d(u,z)-(u-1)l_d(z)}.$$

## 4. A classification for limit laws of random mappings parameters

We begin with a proposition which applies to functional equations of trees.

PROPOSITION 1. *Let $F(u,z,a(u,z))$ be a power series in three variables with non-negative coefficients and $F(0,0,0) = 0$. Suppose that the system of equations $\{\tau = F(1,\rho,\tau), 1 = F'_a(1,\rho,\tau)\}$ has positive solutions $\rho$ and $\tau$ such that $F'_z(1,\rho,\tau) \neq 0$ and $F''_{aa}(1,\rho,\tau) \neq 0$. Then, $F(u,z,a) = 0$ has for solution*

$$a(u,z) = t(u,z) - h(u,z)\sqrt{1-z/\rho(u)},$$

*with $t, h, \rho$ analytic,*

$$t(1,\rho(1)) = \tau(1) \equiv \tau, \quad \rho(1) = \rho \quad \text{and} \quad h(1,\rho(1)) = \sqrt{\frac{2\rho F'_z(1,\rho,\tau)}{F''_{aa}(1,\rho,\tau)}}.$$

We arrive to a general theorem which seems to be the proper theorem to discuss random mappings. We consider a generating function $g(u,z) = \sum_{n,k} g_{n,k} u^k z^n$ corresponding for variables $X_n$ to a probability distribution $\Pr(X_n = k) = g_{n,k}/g_n$. We consider a local expansion in the neighbourhood of $u = 1, z = \rho(u)$, of the form

$$g(u,z) = \frac{1}{1-T(u,z)+h(u,z)\sqrt{1-z/\rho(u)}}.$$

$T, h$ and $\rho$ are analytic and $T(1,\rho) = 1$.

THEOREM 1. *With these hypotheses ($T, h, \rho$ analytic and $T(1,\rho) = 1$),*

1. *If $\rho'(u) = 0$ and $T'_u(1,\rho) > 0$, then $X_n/\sqrt{n} \to \mathcal{R}(\lambda)$, where $\lambda = \frac{1}{2}\left(\frac{h(\rho,1)}{T'_u(\rho,1)}\right)^2$ and $\mathcal{R}(\lambda)$ is the Rayleigh distribution of density $\lambda x \exp\left(-\frac{\lambda}{2}x^2\right)$. Moreover $E(X_n) \approx \sqrt{\frac{\pi n}{2\lambda}}$ and $\mathrm{Var}(X_n) \approx \left(2-\frac{\pi}{2}\right)\frac{n}{\lambda}$.*
2. *If $\rho'(1) \neq 0$ and $T'_u(1,\rho) = 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1)$, where $\mu = -\rho'(1)/\rho(1)$ and $\sigma^2 = \mu^2 + \mu - \rho''(1)/\rho(1)$. Moreover $E(X_n) \approx \mu n$ and $\mathrm{Var}(X_n) \approx \sigma^2 n$.*
3. *If $\rho'(1) \neq 0$ and $T'_u(1,\rho) \neq 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1) \star \mathcal{R}(\sigma^2\lambda)$, where $\mu$ and $\sigma$ are defined as in (2), $\lambda$ is defined as in (1) and the star operator represents the convolution operation.*

REMARK. If $T(1,\rho) \neq 1$, then $\frac{X_n - \mu_n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1)$, (except if $\rho'(u) = 0$ and $T(1,\rho) < 1$, in which case $X_n \to \delta\mathcal{G}$, derivative of a geometric law).

The density and characteristic functions in these different cases are as follows.

1. $\mathcal{R}$ (Rayleigh) $f_{\mathcal{R}(\lambda)}(x) = \lambda x e^{-\lambda x^2/2}$, and $\phi_{\mathcal{R}}(\theta) = 1 + i\theta\sqrt{\frac{\pi}{2}}e^{-\theta^2/2}(1 - i\,\mathrm{erf}(\theta/\sqrt{2}))$.
2. $\mathcal{N}$ (Normal) $f_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\phi_{\mathcal{N}}(\theta) = e^{-\theta^2/2}$.

3. $\mathcal{N} \star \mathcal{R}$ (Normal conv. Rayleigh) $f_{\mathcal{N}\star\mathcal{R}}(x) = (e^{-x^2/4} - e^{-x^2/2})/\sqrt{2\pi} + \frac{xe^{-x^2/4}}{2\sqrt{2}}\mathrm{erf}(x/2)$ and $\phi_{\mathcal{N}\star\mathcal{R}}(\theta) = \phi_{\mathcal{N}}(\theta) \times \phi_{\mathcal{R}}(\theta)$.

PROOF. (Sketch) Let $g(u,z) = \sum_{n\geq 0} p_n(u)z^n/n!$ with $p_n(1) = g_n$. The proof rests on the convergence of the corresponding characteristic functions to (1) $\phi_{\mathcal{R}}(\theta)$, (2) $e^{-\theta^2/2}$, (3) $e^{-\theta^2/2} \times \phi_{\mathcal{R}}(\theta)$. For instance, in case (1), the characteristic function $p_n(e^{i\theta/\sqrt{n}})/g_n$ converges to $\phi_{\mathcal{R}}(\theta)$. The proofs in the different cases make use of Cauchy inversions along suitable contours of the complex plane [1]. □

## 5. Applications

We note $\Xi_n$ the law of $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}}$.

*Leaves.* We have $a(z) = t(u,z) - h(u,z)\sqrt{1 - z/\rho(u)}$. This gives $\{\tau = \rho e^\tau + (u-1)\rho, 1 = \rho e^\tau\}$, which gives $\{t(1,\rho) \equiv \tau(1) = 1, \rho(1) = \rho\}$, and also by differentiation wrt $u$ $\{\tau' = (\rho e^\tau)' + \rho + (u-1)\rho', 0 = (\rho e^\tau)'\}$, these two last equations give $\{\tau'(1) = \rho, \rho'(1) = -\rho^2 \neq 0\}$. This gives for sequences of trees $t(1,\rho) = 1, \rho'(1) \neq 0, t'_u(1,\rho) \neq 0$, and therefore $\Xi_n \to \mathcal{N} \star \mathcal{R}$. This also gives for functional graphs, with $T(u,z) = t(u,z) - (u-1)z$, $T(1,\rho) = 1, \rho'(1) \neq 0, T'_u(1,\rho) = \tau'(1) - \rho = 0$, and therefore $\Xi_n \to \mathcal{N}$.

*Nodes with in-degree r.* As before, $a(z) = t(u,z) - h(u,z)\sqrt{1 - z/\rho(u)}$. We have $\{\tau = \rho e^\tau + \rho(u-1)\frac{\tau^r}{r!}, 1 = \rho e^\tau + \rho(u-1)\frac{\tau^{r-1}}{(r-1)!}\}$. This gives $\tau(1) = 1$ and $\rho(1) = \rho$. By differentiation wrt $u$, we obtain $\tau'(1) = \rho\left(\frac{1}{r!} - \frac{1}{(r-1)!}\right)$ and $\rho'(1) = \frac{-\rho^2}{r!} \neq 0$. For sequences of trees, we get $t(1,\rho) = 1, \rho'(1) \neq 0$ and, if $r \geq 2$, $t'_u(1,\rho) \neq 0$, which implies $\Xi_n \to \mathcal{N} \star \mathcal{R}$. If $r = 1$, the limit law is normal. For functional graphs, we have $T(u,z) = t(u,z) - z(u-1)\left(\frac{a^{r-1}}{(r-1)!} - \frac{a^r}{r!}\right)$. We get $T(1,\rho) = 1, \rho'(1) \neq 0$, and $T'_u(1,\rho) = 0$, which implies that $\Xi_n \to \mathcal{N}$.

*Nodes at distance d from a cycle.* We have $a_d(u,z) = t_d(u,z) - c_d(u,z)\sqrt{1 - ez}$, with $t_0(z) = ug(z)$, $t_d(u,z) = ze^{t_{d-1}(u,z)}$, $c_0(z) = uk(z)$, $c_d(u,z) = t_d(u,z)c_{d-1}(u,z)$. This gives $\rho' = 0, t_d(1,\rho) = 1, t'_d(1,\rho) = 1$. Applying this results to $g(u,z) = 1/(1 - a_d(u,z))$, we get $T(1,\rho) = 1, T'_u(1,\rho) \neq 0, \rho = Cst$, which implies that $\Xi_n \to \mathcal{R}$.

*Nodes with in-degree r.* (Same method.) We have for sequences of Cayley trees $\xi_n \to \mathcal{N} \star \mathcal{R}$, and for functional graphs $\Xi_n \to \mathcal{N}$.

*Nodes at distance $\geq d$ from a leaf.* (Same method.) From a leaf of their own subtree (sequences of Cayley trees), $\Xi_n \to \mathcal{N} \star \mathcal{R}$. In the general case, $\Xi_n \to \mathcal{N}$.

*Nodes at distance d from a leaf.* (Same method.) If the path contains no cyclic edge, $\Xi_n \to \mathcal{R} \star \mathcal{N}$ (except if $d = 1$, in which case $\Xi_n \to \mathcal{N}$). If cyclic edges are allowed, for $d \leq 2$, we have $\Xi_n \to \mathcal{N}$. (Conjecture: this last result is true for all $d$.)

## Bibliography


[1] Drmota (Michael) and Soria (Michèle). – Images and preimages in random mappings. *SIAM Journal on Discrete Mathematics*, vol. 10, n° 2, May 1997, pp. 246–269.

[2] Vitter (Jeffrey Scott) and Flajolet (Philippe). – Analysis of algorithms and data structures. In van Leeuwen (J.) (editor), *Handbook of Theoretical Computer Science*, Chapter 9, pp. 431–524. – North Holland, 1990.

# Patterns in Random Binary Search Trees

*Philippe Flajolet*
Algorithms Project, INRIA Rocquencourt

October 7, 1996

[summary by Michèle Soria]

**Abstract**

In a randomly grown binary search tree (BST) of size $n$, any fixed pattern occurs with a frequency that is on average proportional to $n$. Deviations from the average case are highly unlikely and well quantified by a Gaussian law. Trees with forbidden patterns occur with an exponentially small probability that is characterized in terms of Bessel functions. The results obtained extend to BST's a type of property otherwise known for strings and combinatorial tree models. They apply to paged trees or to Quicksort with halting on short subfiles. As a consequence, various pointer saving strategies for maintaining trees obeying the random BST model can be precisely quantified. The methods used are based on analytic models, especially bivariate generating functions subjected to singularity perturbation asymptotics.

The binary search tree (BST) model is a model of randomness that applies to binary search trees and heap-ordered trees on random data, Quicksort, multidimensional trees, randomised search trees and treaps, as well as to syntax trees that occur naturally in software engineering. Basic properties of random BST have been studied in detail, for example it is well-known that the average value of depth nodes or height is asymptotically logarithmic in the size of the tree (see e.g. [7, 8, 9, 1, 2]). This talk, based on [4], is devoted to characteristics of the shape produced by the BST model.

## 1. Bivariate generating function

The bivariate generating function $F(z, y)$ for the probability $f_{n,k}$ that a random BST of size $n$ has $k$ occurrences of a given pattern $u$ is defined by

$$F(z, y) := \sum_{n,k} f_{n,k} y^k x^n = \sum_t \lambda(t) y^{\omega[t]} z^{|t|},$$

where $\lambda(t)$ is the probability that a given unlabelled tree $t$ is the shape of a random BST of size $|t|$ ($\lambda(t) = \prod_{v \prec t} 1/|v|$, where the product is over all subtrees $v$ of $t$), and $\omega[t] \equiv \omega_u[t]$ denotes the number of occurrences of pattern $u$ as a subtree of the BST $t$.

The recurrence on $\omega[t]$ is obvious: $\omega[t] = [\![t = u]\!] + \omega[t_0] + \omega[t_1]$, where $t_0, t_1$ denote the left and right subtrees of $t$ and where the bracket notation $[\![P]\!]$ is the indicator of $P$ with value 1 if the predicate $P$ is true and 0 otherwise. From this recurrence $F(z, y)$ can be easily shown to satisfy a linear second order differential equation.

LEMMA 1. *The bivariate generating function $F(z, y)$ satisfies the Riccati equation*

$$\frac{\partial}{\partial z} F(z, y) = F^2(z, y) + (y - 1)\lambda(u)|u| z^{|u|-1}, \qquad F(0, y) = 1 \tag{1}$$

Using equation(1), the mean and variance of the number of occurrences of $u$ are obtained by successive differentiation of $F(z,y)$ with respect to $y$, upon setting $y=1$.

THEOREM 1 (Moments of occurrences). *The number $\Omega_n$ of occurrences of a pattern $u$ in a random BST of size $n$ has mean and variance which are asymptotically linear $\mu_n \sim c_1(u)n$, $\sigma_n \sim c_2(u)n$, where $c_1(u)$ and $c_2(u)$ are effectively expressed in terms of $\lambda(u)$ and $|u|$.*

Linearization of the Riccati differential equation (1) gives, in this case, an explicit solution in terms of Bessel functions.

LEMMA 2. *The bivariate generating function of the number of occurrences of pattern $u$ is*

$$F(z,y) = -\frac{w'(z,y)}{w(z,y)}, \qquad w(z,y) = A_m(\Lambda z^{m+1}) - zB_m(\Lambda z^{m+1}),$$

*with $w'(z,y) = w'_z(z,y)$, $\Lambda = |u|\lambda(u)(1-y)$, $m = |u| \geq 2$, and $A_m, B_m$ are normalized Bessel functions of orders $-1/(m+1)$ and $1/(m+1)$ respectively.*

## 2. Asymptotic analysis

The asymptotic behaviour of the bivariate generating function $F(z,y)$ is dictated by the singularities in the main variable $z$ (which are poles since $F(z,y)$ is the quotient of analytic functions), with the auxiliary variable $y$ entering as a parameter. Different regions of values of the auxiliary variable provide different types of information: probability for excluded patterns ($y=0$) or rare occurrences ($y \approx 0$), limiting distribution for the number of patterns, central ($y \approx 1$) as well as local ($|y|=1$). The techniques of proof belong to the class of singularity analysis methods: it is well-known, after Cauchy's inversion formula that the location of polar singularities of a function $f(z) = \sum f_n z^n$ drives the asymptotic form of its coefficients. For bivariate generating function $F(z,y)$, we first perform one level of inversion, resulting in estimates of $f_n(y) = [z^n]F(z,y)$, with an extra property of uniform error bounds for some values of $y$. One more inversion is required in order to recover the individual probabilities $f_{n,k}$. The most direct approach consists in applying *Levy's continuity theorem* for characteristic functions; this implies estimating $f_n(y)$ for $y = e^{i\theta}$, but only $\theta$ near 0 is required because of scaling. Thus, we have a *perturbation* of the univariate problem at $y=1$. When $f_n(y)$ is a "quasi-power", meaning that it behaves very nearly like the powers of a fixed function, a Central Limit Theorem can be derived. A local limit law of the Gaussian type also holds when one can estimate *globally* $f_n(y)$ by quasi-powers in larger regions, like $|y|=1$, and not merely *locally* near $y=1$. In that case, the recovery of $f_{n,k}$ from $f_n(y)$ is achieved by using a saddle-point method.

**2.1. Trees with excluded patterns.** Asymptotic analysis of univariate and bivariate generating functions derived from $F(z,y)$ depends on locating the zeros of $w(z,y)$ where $y$ is a parameter. We thus define the function $\alpha_u(y)$ to be the root of smallest modulus of the Bessel type equation

$$A_m(\Lambda\alpha^{m+1}) - zB_m(\Lambda\alpha^{m+1}) = 0, \qquad \Lambda = (1-y)|u|\lambda(u). \tag{2}$$

This definition specifies $\alpha_u(y)$ unambiguously, for $|1-y|$ not too large ($|1-y| < 5/2$ as is shown in lemma 3 of [4]). The probability that a random BST of size $n$ does not contain the pattern $u$, is by definition $[z^n]F(z,0)$. The following result comes from singularity analysis of meromorphic functions.

THEOREM 2 (Excluded patterns). *The probability $e_{u,n}$ that a random BST of size $n$ does not contain the pattern $u$ satisfies*

$$e_{u,n} = \alpha_u(0)^{-n-1}(1 + O(K^{-n}))$$

*where $K = K(u)$ is a constant strictly larger than 1, and $\alpha_u(0)$ is the smallest positive root of Equation (2) with $y = 0$.*

The same argument proves that, for small enough $|y|$, $[z^n]F(z,y) \sim \alpha_u(y)^{-n-1}$, with a uniform exponentially small error term. The probability that a random BST of size $n$ has $k$ occurrences of a pattern $u$ is obtained by differentiating $k$ times this asymptotic expansion.

THEOREM 3 (Poisson law for rare occurrences). *Given a fixed pattern $u$, for each* fixed *$k$, one has as $n \to \infty$:*

$$\Pr\{\omega_u[t] = k \mid |t| = n\} = \alpha_u(0)^{-n-1} \cdot \frac{(\mu n)^k}{k!}\left(1 + O(\frac{1}{n})\right), \qquad \mu = -\frac{\alpha_u'(0)}{\alpha_u(0)}.$$

**2.2. Gaussian limit laws.** The application of Levy's continuity theorem that leads to a Central Limit Theorem, relies on evaluating $f_n(y)$ when $y = e^{it/\sqrt{n}}$. By the implicit function theorem and the preparation theorem of Weierstrass (see e.g. [6]), there is a small complex neighbourhood of 1 such that the function $\alpha_u(y)$ is analytic and by the analysis of meromorphic functions, $f_n(y)$ is closely approximated by a large power of a fixed function: $f_n(y) = \alpha_u(y)^{-n-1}(1 + O(K^{-n}))$, a situation that conduces to normal laws. The speed of convergence is bounded by means of the Berry-Esseen inequality (see e.g. [3]) that relates the distance between distribution functions and characteristic functions.

THEOREM 4 (Central law for occurrences). *Given a pattern $u$, the number of occurrences $\Omega_n$ of $u$ in a random BST of size $n$ obeys a central limit law with linear mean $\mu_n$ and variance $\sigma_n^2$ (see theorem 1), and the speed of convergence is $O(1/\sqrt{n})$:*

$$\sup_{x \in R}\left|\Pr\left\{\frac{\Omega_n - \mu_n}{\sigma_n} \le x\right\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-w^2/2}\,dw\right| < \frac{1}{\sqrt{n}}.$$

The obtention of a local limit law starts again with the quasi-powers form of $f_n(y)$ and uses a saddle-point derivation.

THEOREM 5 (Local law for occurrences). *Given a pattern $u$, if $|\alpha_u(y)| \neq 1$ for all $y$ with $|y| = 1$, $y \neq 1$, then the random variable $\Omega_n$ satisfies a local limit law:*

$$\sup_x\left|\sigma_n \Pr\{\Omega_n = \lfloor \mu_n + x\sigma_n \rfloor\} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right| \to 0,$$

*for $x$ in any fixed compact subset of $]-\infty, +\infty[$.*

## 3. Other applications

Suitable adaptations of the technique also lead to a distributional analysis of paging and factored representation of BST as a DAG.

**3.1. Paged trees.** Given a tree $t$, its *$b$-index* is a tree that is constructed by retaining only those internal nodes of $t$ which correspond to subtrees of size $> b$. Such an index is well-suited to "paging", where one has a two-level hierarchical memory structure: the index resides in main memory and the rest of the tree is kept in pages of capacity $b$ on peripheral storage. Let $\iota[t] = \iota_b[t]$ denote the size of the $b$-index of $t$. The analysis is then clearly equivalent to determining the total number of occurrences of all patterns of size $\le b$. The bivariate generating function $G(z,y) := \sum_t \lambda(t) y^{\iota[t]} z^{|t|}$ satisfies a Riccati equation

$$\frac{\partial}{\partial z}G(z,y) = yG^2(z,y) + (1-y)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right),$$

which is transformed by linearization. This equation has no explicit solution, but the use of Weierstrass Preparation theorem shows that $G(z, y)$ has a unique dominant simple pole for $y$ in a small neighbourhood of 1. Thus $[z^n]G(z, y)$ reduces to quasi-powers, and a central limit law follows:

THEOREM 6 (Paging distribution). *For fixed $b \geq 2$, the size $I_n$ of the $b$-index constructed on a random BST of size $n$ has mean $\mu_n$ and variance $\sigma_n^2$ that satisfy*

$$\mu_n = \frac{2(n+1)}{b+2} - 1, \qquad \sigma_n^2 = \frac{2}{3}\frac{(b-1)b(b+1)}{(b+2)^2}(n+1).$$

*The random variable $I_n$ obeys a central limit law with speed of convergence $O(1/\sqrt{n})$.*

**3.2. Factored representations of trees.** We consider finally a global parameter $\kappa[t]$ of trees that represents the number of structurally different subtrees (*i.e.*, number of different subtree shapes) that occur in $t$. This parameter is of intrinsic interest as an indicator of the structural "richness" of $t$. It also measures the optimal storage complexity of tree $t$ when all common subtrees are factored and represented only once. Then, $\kappa[t]$ measures the number of nodes of the maximally factored DAG (directed acyclic graph) corresponding to $t$, a quantity that intervenes in parsing and data compression applications [5].

The quantity that is actually analysed is the size of a DAG representation which is partly redundant (all trees of size less than $b$ are represented once, irrespective of their possible nonoccurrence in $t$) and partially factored (nodes commanding subtrees of size $\geq b$ are each represented irrespective of the fact that they may be associated to repeated subtrees).

THEOREM 7. *The average value of the DAG size of a random BST of size $n$ satisfies the upper bound,*

$$K_n \leq 4 \log 2 \frac{n}{\log n} + O\left(\frac{n \log\log n}{(\log n)^2}\right).$$

This upper bound on $K_n$ is of the right order, since L. Devroye has shown a lower bound in $O(n/\log n)$ (unpublished paper, May 97); but the constant is still unknown.

## Bibliography


[1] Devroye (Luc). – A note on the expected height of binary search trees. *Journal of the ACM*, vol. 33, n° 3, 1986, pp. 489–498.

[2] Devroye (Luc). – Limit laws for local counters in random binary search trees. *Random Structures and Algorithms*, vol. 2, n° 3, 1991, pp. 303–315.

[3] Feller (William). – *An Introduction to Probability Theory and Its Applications.* – John Wiley & Sons, New York, 1971, 2nd edition, vol. II.

[4] Flajolet (Philippe), Gourdon (Xavier), and Martínez (Conrado). – *Patterns in random binary search trees.* – Research Report n° 2997, Institut National de Recherche en Informatique et en Automatique, October 1996. 23 pages. To appear in *Random Structures & Algorithms.*

[5] Flajolet (Philippe), Sipala (Paolo), and Steyaert (Jean-Marc). – Analytic variations on the common subexpression problem. In Paterson (M. S.) (editor), *Automata, Languages and Programming, Lecture Notes in Computer Science*, vol. 443, pp. 220–234. – 1990. Proceedings of the 17th ICALP Conference, Coventry, July 1990.

[6] Hille (Einar). – *Analytic function theory. Vol. 1.* – Ginn and Company, Boston, 1959, xi+308p. Introduction to Higher Mathematics.

[7] Knuth (Donald E.). – *The Art of Computer Programming.* – Addison-Wesley, 1973, vol. 3: Sorting and Searching.

[8] Mahmoud (Hosam M.). – *Evolution of Random Search Trees.* – John Wiley & Sons Inc., New York, 1992, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, xii+324p.

[9] Sedgewick (Robert) and Flajolet (Philippe). – *An Introduction to the Analysis of Algorithms.* – Addison-Wesley Publishing Company, 1996, 512p.

# On the Height Concentration of Binary Search Trees

*Mike Robson*
LaBRI, Bordeaux

June 2, 1997

**Abstract**

The height of a binary search trees with $n$ nodes is equal to the size of the stack used when sorting $n$ elements by the simplest version of Quicksort.

The average value of this height is known $(4.31107... \log(n) + O(\log \log n))$. Its variance is only known to be upper bounded by $O(\log \log n^2)$. It has been conjectured that it is actually $O(1)$.

After a historical survey on this problem, this talks presents new results which make this conjecture more plausible, namely bounds of order $O(1)$ on the absolute difference between the height and its average.

# Randomized Binary Search Trees

*Conrado Martínez*
Universitat Politecnica de Catalunya, Spain

December 9, 1996

[summary by Danièle Gardy]

**Abstract**

Results on the performance of binary search trees are usually relative to the random permutation model. This talk presents a randomized version of the insertion and deletion operations on BST that ensures that, whatever the initial distribution of keys, results from the random permutation model apply.

## 1. Classical and randomized binary search trees

The usual model to analyze the performance of a binary search tree (BST) built on $n$ keys assumes that all possible permutations are equally likely. This ensures that balanced trees have a "large" probability of appearing (a recent paper by Fill [2] gives a mathematical justification of this easy-to-understand fact), and that degenerate trees have a low probability; as a result it is well known that the search, insert or delete operations have expected time proportional to $\log n$. But the assumption that the entries are inserted in random order does not always hold; moreover it is known that, under a specific sequence of insertions and deletions, the resulting tree is no longer random. Hence degeneracy of a BST is a real issue. Martínez and Roura propose here randomized versions of the insertion and deletion operations that ensure that, whatever the sequence of operations and the order of insertions and deletions of keys, the resulting binary search tree has the same probability as if it had been built on the random permutation model. As a consequence, the expected cost of operations on a randomized BST is guaranteed to be of order $O(\log n)$. Their work was first presented in [5]; a detailed version appears in [6].

**1.1. Randomized insertion.** The main idea is to use the *split* operation: This operation creates two BST's from a BST and a key, and can be used if one wishes to insert a key at the root of a BST. This algorithm is presented for example in [3, pp. 202–204]; it works by building two new BST's $\mathcal{L}$ and $\mathcal{R}$ from a key $X$ and a given BST $\mathcal{A}$, which is destroyed by the operation: The first BST $\mathcal{L}$ contains the keys of $\mathcal{A}$ that are smaller than or equal to $X$, and the second BST $\mathcal{R}$ contains the keys of $\mathcal{A}$ that are larger than $X$. Then we insert $X$ into a new node, whose left and right sons are $\mathcal{L}$ and $\mathcal{R}$. An example is given in Figure 1.

The randomized insertion works as follows

> To insert a key $X$ into a BST with $n-1$ keys, we use the *insertion at root* algorithm with probability $1/n$, and with probability $1-1/n$ we insert the key recursively into the right or left subtree, according to the respective values of the key and the root.

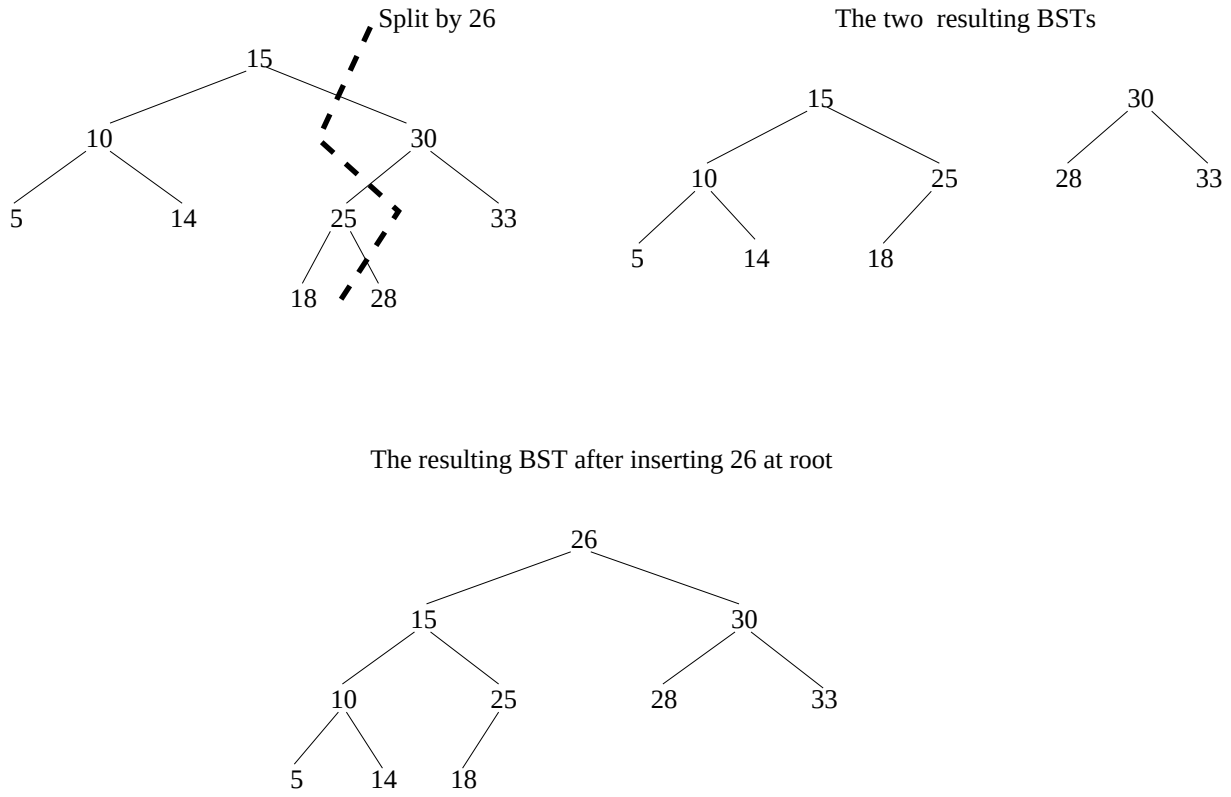



FIGURE 1.

The recursive insertion in a subtree is itself the randomized version; as a consequence, an insertion can happen at any place in the path from the root to the leaf where the key would be inserted if standard insertion were used.

**1.2. Randomized deletion.** As in the "classical" case without randomization, the deletion first searches for the key $X$ to be deleted; the difference appears when deleting the key at the root of a tree. The deletion of the root is done here by the *join* algorithm of Martínez and Roura, which can be summarized as follows:

- If either the left or right subtree of $X$ is empty, just send back the other subtree.
- Otherwise, let $\mathcal{L}$ and $\mathcal{R}$ be the left and right subtrees, of respective sizes $m$ and $n$. Assume that $\mathcal{L} = (a, L_l, L_r)$ and $\mathcal{R} = (b, R_l, R_r)$. Then, with probability $m/(m+n)$, return the BST $(a, L_l, join(L_r, \mathcal{R}))$, and return the BST $(b, join(\mathcal{L}, R_l), R_r)$ with probability $n/(m+n)$.

An example is given in Figure 2.

## 2. Performance analysis

For this analysis, we need to make precise the notion of *random BST*:

> A BST on $n$ keys is random if either it is empty ($n = 0$), or the probability that a given key is at the root is $1/n$, and the left and right subtrees are random.

This is exactly the BST built under the usual permutation model. The main point in the analysis of randomized BST is to show that insertions or deletions always yield a random BST if applied to a random BST, whatever the key to be inserted or deleted. Once this is done, results on random

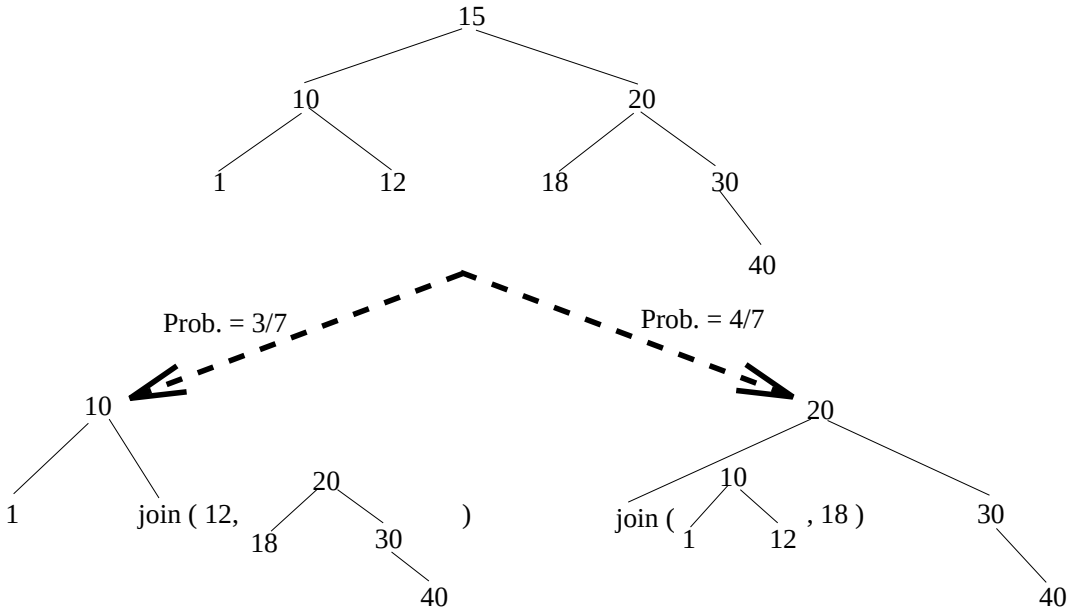



FIGURE 2.

BST's apply: the average cost of an operation on a randomized BST is $O(\log n)$, for any sequence of operations on the tree.

## 3. Storage requirements

The randomized insertion and deletion both require storing in each node of the BST the size of the subtree rooted at this node, hence a memory requirement of order $n$. As a consequence of storing the number of nodes, randomized BST can also be used for *ranking algorithms*.

Each of the algorithms has an iterative version that uses only a constant number of auxiliary variables (no stack). The top-down implementation of the deletion algorithm requires that the size of each subtree be modified when the root is traversed; this can lead to problems if we try to delete a key that is not present in the tree. To avoid this, Martínez and Roura introduce an alternative scheme where each node stores the size of one of its subtrees (any of them) and an "orientation" bit that indicates the subtree whose size is kept; they also keep a single global variable with the total number of keys in the tree.

## 4. Conclusion

This elegant work solves the problem of degeneracy in BST, although the worst-case performance is still linear. Compared to balanced trees, algorithms for randomized BST are notably simpler; their performance can be compared with the behaviour of skip lists ([7], see also for example [1, 4] for the analysis of their performances).

The structure most closely related to randomized BST may well be the *treap* of Aragon and Seidel [8]: a BST is augmented with a priority for each key, in such a way that the tree is a search tree for keys, and a heap for priorities (a node has a larger priority than its sons). When priorities are chosen randomly, it is easy to see that an insertion, for example, can happen at any place on the path from the root to the leaf that would receive the new key in a BST with standard insertion at leaves. A main difference with the present work is that, while the randomness is related to the

tree itself in the work of Martínez and Roura, in treaps it comes from the randomly chosen priority, which may appear less natural.

## Bibliography


[1] Devroye (L.). – A limit theorem for random skip lists. *Annals of Applied Probability*, vol. 2, n° 3, 1992, pp. 597–609.

[2] Fill (James Allen). – On the distribution of binary search trees under the random permutation model. *Random Structures and Algorithms*, vol. 8, n° 1, 1996, pp. 1–25.

[3] Froidevaux (Christine), Gaudel (Marie-Claude), and Soria (Michèle). – *Types de données et algorithmes.* – McGraw-Hill, Paris, 1990.

[4] Kirschenhofer (P.) and Prodinger (H.). – The path length of random skip lists. *Acta Informatica*, vol. 31, n° 8, 1994, pp. 775–792.

[5] Martínez (C.) and Roura (S.). – Randomization of search trees by subtree size. In Díaz (J.) and Serna (M.) (editors), *Proceedings of the 4th European Symposium on Algorithms (ESA). Lecture Notes in Computer Science*, vol. 1136, pp. 91–106. – Springer-Verlag, 1996.

[6] Martínez (C.) and Roura (S.). – *Randomized binary search trees.* – Technical Report n° LSI-97-8, Departament de Llenguatges i Sistemes Informàtics, Universitat Politecnica de Catalunya, Barcelona, 1997.

[7] Pugh (William). – Skip lists: a probabilistic alternative to balanced trees. In Dehne (F.), Sack (J.-R.), and Santoro (N.) (editors), *Algorithms and data structures. Lecture Notes in Computer Science*, vol. 382, pp. 437–449. – Berlin, 1989. Proceedings of a Workshop held at Ottawa, 1989.

[8] Seidel (R.) and Aragon (C. R.). – Randomized search trees. *Algorithmica*, vol. 16, n° 4-5, 1996, pp. 464–497.

# The Analysis of Multiple Quickselect

*Helmut Prodinger*
Technical University of Vienna

June 30, 1997

[summary by Philippe Flajolet]

**Abstract**

Quickselect is a version of Quicksort that makes it possible to find efficiently any element in an unsorted file given its rank. "Multiple Quickselect" is designed to find simultaneously a collection of elements, also specified by their ranks. This talk shows how to analyse Multiple Quickselect when the underlying permutation is random and the collection of ranks is a random $p$-subset ($p$ fixed). The analysis provides a nice illustration of the use of multivariate generating functions.

## 1. Algorithms

The principle of *Quicksort* is of course extremely well known. Given an array $T[1 \dots n]$ of numbers to be sorted, choose a "pivot", say $T[1]$, then partition (by a linear scan) the original array into the two subarrays, $T_<$ and $T_>$, formed by elements that are respectively smaller and larger than the pivot, and proceed recursively. The algorithm was first proposed by Hoare in 1962. Its characteristics are fairly well understood: the algorithms sorts in place (it uses $O(1)$ auxiliary memory in addition to the recursion stack), its average number of comparisons is $\sim 2n \log n$, while the implementation constants are especially low. For these reasons, Quicksort is a method of choice amongst sorting algorithms, and it has been adopted as the basis of the Unix `sort` command. Note however that it is still unknown whether the limit distribution of the number of comparisons that has been proved to exist can be characterized in terms of standard special functions of analysis. (Existence proofs comprise the martingale argument of Régnier, the moment method of Hennequin, and the contraction metrics approach of Rösler; see [4] and references therein.)

*Quickselect* is a simplified version of Quicksort adapted so as to locate the $j$th ranked element in a file. In that case, it suffices to effect one recursive descent in one of the two subfiles $T_<, T_>$ created by the basic partitioning stage of Quicksort. Knuth's book [2] provides the analysis of the two parameters of interest, the average number of passes (recursive calls) $P[n; j]$ and the average number of comparisons $C[n; j]$,

$$P[n; j] = H_j + H_{n+1-j} - 1,$$

$$C[n; j] = 2\Big(n + 3 + (n + 1)H_n - (j + 2)H_j - (n + 3 - j)H_{n+1-j}\Big)$$

where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ denotes the $n$th harmonic number. In particular, the mean number of comparisons is $O(n)$, uniformly for all $j$. From the complexity standpoint, the algorithm thus has the same type of cost as a fixed number of scans of the input file—a highly appealing feature. (Contrast the simplicity of Quickselect with algorithms like median-finding that are constructed specifically to achieve linear complexity in the worst-case!)

*Multiple Quickselect* is an extension of Hoare's idea that works as follows: Assume that we simultaneously search for $p$ elements with ranks $J = (j_1, j_2, \ldots, j_p)$ where $1 \le j_1 < j_2 < \cdots < j_p \le n$ is a fixed set of $p$ values. Then, depending on the interval determined by $j_1, \ldots, j_p$ in which the pivot element falls, we continue recursively on both subfiles $T_<, T_>$, but with smaller $J$-index sets. The principle is quite simple and a nice description of the algorithm can be found in [3]. Notice that Multiple Quickselect can be used for instance to determine quantiles of distributions efficiently.

In a previous work, Prodinger [7] has given explicit formulæ for both the average number of passes $P[n; j_1, \ldots, j_p]$ and the average number of comparisons, $C[n; j_1, \ldots, j_p]$. As a consequence, the so-called "*grand averages*",

$$\mathcal{P}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le n} P[n; j_1, \ldots, j_p],$$

$$\mathcal{C}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le n} C[n; j_1, \ldots, j_p]$$

have been determined (see also [3]). However, the approach used in [7] was a mixture of guessing and proofs by induction.

The present paper by Panholzer and Prodinger [6] develops a direct generating function approach to the analysis of the grand averages. This gives access to variances that were out of reach with the old method.

## 2. Moments

In both instances, namely number of passes and of comparisons, the problem is modelled by a trivariate generating function $\Phi(z, u, v)$, where the coefficient $n![z^n u^p v^m]\Phi$ is the number of permutations of $\{1, 2, \ldots, n\}$ and of subsets of $p$ elements of $\{1, 2, \ldots, n\}$ such that the parameter of interest has value $m$. Under the random permutation model, the splitting probabilities of Quicksort and Quickselect are given by

$$\Pr\{K = k\} = \frac{1}{n},$$

where the random variable $K \in [1, n]$ denotes the rank of the pivot. The model is thus isomorphic to that of binary search trees (BST's) or heap-ordered trees [2, 4].

*Number of passes.* The trivariate generating function (GF) where $z$ records the size $n$ of the input permutation, $v$ records the number of passes (recursive calls), and $u$ records the number of elements being simultaneously selected satisfies

$$\Phi'(z, u, v) = v(1 + u)\Phi^2(z, u, v) + \frac{1 - v}{(1 - z)^2} \tag{1}$$

with $\Phi(0, u, v) = 1$ and $\Phi'$ denotes a partial derivative with respect to $z$. The relation (1) reads off almost directly from the problem. We may also view it as expressing the size of a generalized common ancestor tree in binary search trees.

As is usual in distributional analyses of additive parameters on BST's, we have a Riccati equation; see for instance, the analysis of patterns in BST's in [1]. Such equations are systematically linearized by a change of variable of the form $\Phi(z) = a(z)W'(z)/W(z)$. Here,

$$\Phi(z, u, v) = \frac{\Omega + 1 - 2v + (1 - z)^{\Omega}(\Omega - 1 + 2v)}{\Big(\Omega + 1 - 2v(1 + u) + (1 - z)^{\Omega}(\Omega - 1 + 2v(1 + u))\Big)(1 - z)} \tag{2}$$

$$\Omega = \sqrt{1 - 4v(1 + u)(1 - v)}.$$

(This is well within the automatic capabilities of Maple.) By differentiating with respect to $v$ and expanding, we obtain a result of [7].

THEOREM 1. *For $p \geq 1$ the average number $\mathcal{P}_{n,p}$ of passes employed by Multiple Quickselect in order to search for $p$ random elements is*

$$\mathcal{P}_{n,p} = (H_{n+1} - H_p)\,\frac{2p(n+1)^2}{(n+2-p)(n+1-p)} - \frac{n(2p-1)+p}{n+2-p}. \tag{3}$$

A particular case is the basic Quickselect algorithm ($p = 1$), for which

$$\mathcal{P}_{n,1} = 2\left(1+\frac{1}{n}\right)H_n - 3.$$

By taking second derivatives, we have access to the variance. A careful expansion guided by educated guesses and patience then gives explicit expressions. We state the result from [6] as it is a good indication of the intricacies involved.

THEOREM 2. *The variance of the number of passes when searching for a random set of $p \geq 2$ elements with Multiple Quickselect is given by*

$$\begin{aligned} V_{n,p} = &-\frac{4p(n+1)^2\Psi_1}{(n+4-p)(n+3-p)(n+2-p)^2(n+1-p)^2}\Big(H_{n+1}-H_p\Big)^2 \\ &+\frac{2(n+1)^2\Psi_2}{(n+4-p)(n+3-p)(n+2-p)^2(n+1-p)}\Big(H_{n+1}-H_p\Big) \\ &-\frac{4p(n+2)(n+1)^2(pn+p+2)}{(n+4-p)(n+3-p)(n+2-p)(n+1-p)}\Big(H^{(2)}_{n+1}-H^{(2)}_p\Big) \\ &+\frac{2(n+1)^2\Psi_3}{(n+4-p)(n+3-p)(n+2-p)^2} \end{aligned}$$

$$\begin{aligned} \Psi_1 &= (3p-2)n^3 - 2(p^2-9p+5)n^2 - (p^3+5p^2-33p+16)n - p^3 - 5p^2 + 20p - 8 \\ \Psi_2 &= pn^3 + (11p^2-13p+8)n^2 - (9p^3-54p^2+62p-32)n - 3p^4 - p^3 + 38p^2 - 56p + 32 \\ \Psi_3 &= (2p-1)n^2 - 2(5p^2-15p+8)n + 8p^3 - 45p^2 + 72p - 32. \end{aligned}$$

For $p = 1$, the variance simplifies to

$$V_{n,1} = -4\frac{n+1}{n^2}H^2_{n+1} + \frac{2(n+4)(n+1)}{n^2}H_{n+1} - 4\frac{n+1}{n}H^{(2)}_{n+1} + \frac{2(2n^2-n-2)}{n^2}.$$

*Number of comparisons.* The process for moment calculations is similar, only the expressions become a little more complicated. The fundamental functional equation is now of the difference-differential type,

$$\Phi'(z,u,v) = (1+u)\Phi^2(zv,u,v) + \frac{1}{(1-z)^2} - \frac{1}{(1-zv)^2},$$

with $\Phi(0,u,v) = 1$. Contrary to the previous case, this equation is not known to have a closed form solution. It is however still possible to "pump" GF's of moments by differentiation. We must omit details here and simply state the average-case result while referring to [6] for a daunting variance computation that involves the dilogarithm $\mathrm{Li}_2(z) := \sum z^n/n^2$.

THEOREM 3. *For $p \geq 1$ the average number $\mathcal{C}_{n,p}$ of comparisons to search for $p$ random elements with Multiple Quickselect is given by*

$$\mathcal{C}_{n,p} = -\frac{2p(n+1)(4n-p+5)}{(n+2-p)(n+1-p)}H_{n+1} + \frac{2(n+1)^2(n+2p+2)}{(n+2-p)(n+1-p)}H_p + \frac{n^2+(5p-1)n+3p}{n+2-p}.$$

In particular for a basic (single element) Quickselect and a full sort, we get back the classical results,

$$\mathcal{C}_{n,1} = 3n - 8\left(1+\frac{1}{n}\right)H_n + 13, \qquad \mathcal{C}_{n,n} = 2(n+1)H_n - 4n.$$

## 3. Distributions

The limiting distribution of Quicksort has been under attack for about 20 years. For basic Quickselect, the problem is in a way simpler since the recursion structure is a linear one. Here is what is known at present:

- The number of passes of multiple quickselect ($p$ fixed) is asymptotically normal. This observation results rather directly from singularity perturbation methods applied to the explicit form (2) of the trivariate GF (Flajolet & Prodinger 1997, unpublished).
- The number of comparisons of basic (single) Quickselect has been studied by Mahmoud *et al.* [5] who determined explicitly the characteristic function of the limit law. The main result of [5] entails that this limit has the same distribution as the sum of a Poisson number of independent random variables with an elementary density.

The second result suggests that there is an interesting class of limit distributions for comparison costs when $p$ is a fixed integer $p \geq 2$.

## Bibliography


[1] Flajolet (Philippe), Gourdon (Xavier), and Martínez (Conrado). – *Patterns in random binary search trees.* – Research Report n° 2997, Institut National de Recherche en Informatique et en Automatique, October 1996. 23 pages. To appear in *Random Structures & Algorithms.*

[2] Knuth (Donald E.). – *The Art of Computer Programming.* – Addison-Wesley, 1973, vol. 3: Sorting and Searching.

[3] Lent (Janice) and Mahmoud (Hosam M.). – Average-case analysis of multiple Quickselect: an algorithm for finding order statistics. *Statistics & Probability Letters*, vol. 28, n° 4, 1996, pp. 299–310.

[4] Mahmoud (Hosam M.). – *Evolution of Random Search Trees.* – John Wiley & Sons Inc., New York, 1992, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, xii+324p.

[5] Mahmoud (Hosam M.), Modarres (Reza), and Smythe (Robert T.). – Analysis of QUICKSELECT: an algorithm for order statistics. *RAIRO Theoretical Informatics and Applications*, vol. 29, n° 4, 1995, pp. 255–276.

[6] Panholzer (Alois) and Prodinger (Helmut). – A generating functions approach for the analysis of grand averages for multiple quickselect. – Preprint, 1997.

[7] Prodinger (H.). – Multiple quickselect — Hoare's find algorithm for several elements. *Information Processing Letters*, vol. 56, 1995, pp. 123–129.

# Towards Analytical Information Theory: Recent Results on Lempel-Ziv Data Compression Schemes

*Wojciech Szpankowski*
Purdue University

September 23, 1996

[summary by Philippe Flajolet]

**Abstract**

The Lempel-Ziv algorithms are well-known dynamic dictionary algorithms of use in data compression. The talk shows how analytic models originally developed for the analysis of tries and digital search trees may be used to characterize their compression characteristics.

## 1. The Lempel-Ziv algorithms

A *text* to be compressed is always a message to be transmitted (to your friend, to your laser printer) that has a "meaning", hence a certain structure that goes along with some sort of "redundancy". In a natural language like English, trigrams like 'ted' or 'ing' are much more likely to be encountered, and much more frequently so, than 'qrm' or 'bzw', and a text on data compression is likely to contain more repetitions of 'algorithms' and fewer of 'smoking' than a text on public health. Compression algorithms precisely try to capture such regularities.

These observations have given rise to a first generation of methods. For a text in English, list all the conceivable trigrams (say!) in the language and transmit trigrams codes instead of individual letters. In addition, shorter codes may be assigned to more frequent trigrams, yielding further gains—this can be done efficiently by Huffman's algorithm. These methods are known as "static dictionary algorithms". They have the obvious drawback of not being adaptive; a scheme originally designed for English text is not likely to accommodate well Sanskrit epics, postscript code, or image bitmaps.

Such was essentially the state of the art before the appearance of the celebrated Lempel-Ziv (LZ) papers in 1977 and 1978; see [10, 11]. The LZ algorithms—there are two basic ones and a denumerable collection of variants—can be viewed as building adaptively a "dynamic dictionary" that is dependent upon the particular text subjected to compression. Their common basis is:

THE "DEJA VU" PRINCIPLE.

(Pronounce as "day-jah voo"!) As the text proceeds, it is parsed into *segments* also called *phrases*. Instead of transmitting the letters of the text itself, just transmit *references* to places in the text where each segment was encountered before.

More precisely, the LZ77 and LZ78 algorithms, as considered here, are defined by:

**LZ77.** Scan the text. Starting a new segment, search for the longest matching factor already encountered in the past. Transmit its reference (position and length) and the next letter.

**LZ78.** Scan the text. Starting a new segment, search for the longest matching segment already encountered in the past. Transmit its reference (rank) and the next letter.

For instance, given a text that is a long sequence of a's, the two parsings start like

LZ77: | a | aa | aaaa | aaaaaaaa | ···
LZ78: | a | aa | aaa | aaaa | aaaaa | ···

In other words, LZ77 defines the new segment as anything (of maximal length) that has already been seen before, possibly across boundaries of previously defined segments, while LZ78 respects the boundaries of previously defined segments.

For LZ77, the transmission of each segment is of the type `<position, length, letter>` (position of the previous occurrence and length of the factor, plus new letter). For LZ78, it is of the type `<rank, letter>` (rank of the already encountered segment, plus new letter). So, in the case of our long string of a's, the next segment formed is encoded as

LZ77: `<1, 15, a>`, LZ78: `<5, a>`.

For LZ77, this says: "repeat the 15 characters starting at position 1 in the text and append an a". For LZ78: "repeat the 5th segment already parsed and append an a". Reconstruction of the source text at the receiving end is then particularly easy as it suffices to "expand" the references.

Another example is provided by the two sequences

LZ77: `<0,0,a> <0,0,b> <0,0,r> <1,1,c> <1,1,d> <1,4,#>`
LZ78: `<0,a> <0,b> <0,r> <1,c> <1,d> <1,b> <3,a> <0,#>`

that encode 'abracadabra# (with a terminator symbol). As these examples demonstrate, the LZ77 algorithm "learns" faster but, the implied dictionary being larger, references are more costly as their encodings require more bits. Fine analysis is thus needed in order to characterize the various tradeoffs involved.

*Variants and implementations.* There exist a great many variants of the LZ77 and LZ78 algorithms.

**LZ77:** One can consider that the entire alphabet sequence is prepended to the text. In this case, the new segments need not include a new-letter field, as each letter at least has been already encountered before. Also, one can limit the past fraction of the text to which the "deja vu" principle is applied (historically, to 8192 characters). One obtains in this way an algorithm close to the original LZ77 specification.

**LZ78:** Similarly, one can update the dictionary by deleting old segments (say, on a least recently used basis), and thus maintain a dictionary of an *a priori* bounded size.

In **LZ77**, when a new segment is started at position $(n+1)$ in the text, one must look for a longest factor that has occurred before. This corresponds to a virtual dictionary $\mathcal{D}$ that should contain all the $\frac{1}{2}n(n+1)$ factors of the text, as seen so far. There are two immediate solutions to this problem: $(i)$ don't store the dictionary and use naïve string searching in the text itself; $(ii)$ build a digital tree (also known as *trie*) of all suffixes of the part of the text seen so far. The time/space complexity pairs of these solutions are $\langle O(n^2), O(n)\rangle$ and $\langle O(n^2), O(n^2)\rangle$, respectively. The trie solution is interesting since a construction (the *suffix* tree) is known in order to eliminate duplication of informations; the suffix tree implementation has an $\langle O(n), O(n)\rangle$ complexity but is somewhat intricate to implement.

In contrast, **LZ78** is much easier to implement. It suffices to maintain a digital search tree of all the segments encountered so far. A new segment is detected by following a path in the tree. Each internal node of the tree is associated with such a segment seen in the past and insertion takes place at an external node, which corresponds to extending a previously encountered segment.

These algorithms are famous. They are used in the Unix `compress` and `zip` commands (based on LZ78 and LZ77, resp.), in data transmission (the V42bis standard for modems), in image compression (the `gif` format), etc. Jump to your favourite web search engine for details.

## 2. Models and analysis

Classical information theory teaches us that the best rate at which a text can be compressed is given by the *entropy* function. Here, we concentrate on a binary memoryless channel, where each character in the source text has probability $p$ of being a 0 and probability $q = 1 - p$ of being a 1. In that case, the (binary) entropy is

$$h = p \log_2 \frac{1}{p} + q \log_2 \frac{1}{q}. \tag{1}$$

A random text of $n$ characters can then be compressed into *no less* than

$$h \cdot n(1 + o(1)) \tag{2}$$

bits, on average. The *redundancy* of a compression scheme is a measure of its distance to the information-theoretic lower bound (2).

It is already known that the two Lempel-Ziv algorithms achieve asymptotically the lower bound, so that the redundancy per character is $o(1)$ for both schemes. In other words, what is needed for further comparison is a *second order* asymptotic analysis of these algorithms. The talk centers on the LZ78 algorithm that has a mathematically pleasant decomposable structure closely related to *digital search trees*. Digital search tree intervene both as a data structure in the implementation of LZ78 and as a probabilistic model of random trees.

*Digital search trees.* Recall that a *digital search tree* or DST is a hybrid of the digital trie and the binary search tree defined as follows. A *sequence* $S = (s_1, \ldots, s_m)$ of $m$ binary strings is given. The digital search tree $\mathrm{dst}(s)$ is recursively defined by: $(i)$ the root contains the first string $s_1$; $(ii)$ the left subtree is formed by taking the subsequence $S^{[0]}$ of $(s_2, \ldots, s_m)$ of strings starting with a 0 and stripped of this initial 0; $(iii)$ the right subtree is formed similarly from strings starting with a 1. In other words,

$$\mathrm{dst}(S) = \langle s_1, \mathrm{dst}(S^{[0]} \backslash 0), \mathrm{dst}(S^{[1]} \backslash 1) \rangle,$$

where $S \backslash j$ means the sequence $S$ with all its elements stripped of their initial letter $j$, and with $\mathrm{dst}(s) = \langle s, \emptyset, \emptyset \rangle$ for a one element sequence.

Assume that strings obey the Bernoulli model where each component bit $b$ satisfies

$$\Pr\{b = 0\} = p, \qquad \Pr\{b = 1\} = q = 1 - p,$$

independently of the others. This model implies the *random DST model* [9], where a tree of size $m$ has a left subtree of size $K$ and a right subtree of size $m - 1 - K$ satisfying

$$\Pr\{K = k\} = \binom{m-1}{k} p^k q^{m-1-k}. \tag{3}$$

In the unbiased case, $p = q = \frac{1}{2}$, this model has been analysed by Knuth, Coffman and Eve, as well as Konheim and Newman; see the references in [9]. It is known for instance that the expected path length, under the biased model, is of the form

$$m \log_2 m + m \left( \frac{\gamma - 1}{\log 2} + \frac{3}{2} - \alpha + \delta(\log_2 m) \right) + O(m).$$

There $\alpha := \sum_{k \geq 1} (2^k - 1)^{-1}$ and $\delta$ is a periodic function with magnitude less than $10^{-6}$. Height and other parameters are studied by Aldous and Shields in [1].

*Data compression.* Consider an execution of the LZ78 algorithm when $n$ characters of the text have been scanned, assuming that a new segment starts at position $(n+1)$. Under the binary memoryless model, the number of segments formed so far is a random variable $M_n$ that is also the number of internal nodes of the companion dst built by the algorithm. By design, the dst is built of keys that obey Eq. (3), with $M_n = m$. It is also realized easily that the internal path length of the corresponding tree is equal to $n$.

Thus, there are two closely related models that are in a way "inverse" of each other:

- the *random dst model* where the number of strings $m$ is fixed, and path length of the tree is to be analysed;
- the *compression model* where a dst is built by successive insertions until path length attains a fixed bound $n$, and the size of the tree (corresponding to the size of the compressed text) is to be analysed.

Analysis starts with the random dst model. Inversion will then be realized by an application of renewal theory.

## 3. The random digital search tree model

From an analytical standpoint, this model is the most natural one, since the random tree process given by (3) is a decomposable one. The novelty comes from the fact that, till recently, only the unbiased case $p = q = \frac{1}{2}$ had been analyzed. Jacquet and Szpankowski [7] have proved the following.

THEOREM 1. *Let $L_m$ be the path length of a digital search tree built on $m$ random strings according to the Bernoulli model. Then, the mean and variance of $L_m$ satisfy*

$$\mathrm{E}\, L_m = \frac{m}{h}\left(\log_2 m + \frac{h_2}{2h} + \gamma - 1 - \alpha + \delta_0(\log m)\right) + O(\log m),$$

$$\mathrm{Var}\, L_m = c_2 m \log m + O(m),$$

*with $\delta_0$ a periodic function of mean value 0 and small amplitude, and*

$$h = p\log_2\frac{1}{p} + q\log_2\frac{1}{q},\ h_2 = p\log_2^2 p + q\log_2^2 q,\ \alpha = -\sum_{k=1}^{\infty}\frac{p^{k+1}\log_2 p + q^{q+1}\log_2 q}{1-p^{k+1}-q^{k+1}},\ c_2 = \frac{h_2 - h^2}{h^3}.$$

*Furthermore, the distribution of $L_m$ is asymptotically normal,*

$$\frac{L_m - \mathrm{E}\, L_m}{\sqrt{\mathrm{Var}\, L_m}} \xrightarrow{d} \mathcal{N}(0,1).$$

As is well-known, there are three ways to conduct such an analysis, with Poisson, exponential, or ordinary generating functions (PGF, EGF, OGF). The proof given in [7] is of the Poisson type. The bivariate PGF satisfies then the nonlinear difference-differential equation,

$$\frac{\partial L(z,u)}{\partial z} = L(pzu,u)L(qzu,u),$$

with $L(z,0) = 1$. The approach starts with a quasi-linearization technique of [5], by considering $\ell(z,u) = \log L(z,u)$. By bootstrapping, the functional equation is solved in larger and larger pseudo-cones. Solutions are estimated asymptotically by means of the Mellin transform technology summarized in [3]. Then, mean, variance, and limit distribution follow for the Poisson model. Translation to the Bernoulli model (*i.e.*, the binary memoryless channel) is then achieved by "analytic depoissonization", *i.e.*, an adequate use of the saddle-point method originating in Jacquet and Régnier's analysis of path length in tries. The technical difficulties are rather formidable, but the results obtained are extremely precise.

## 4. The compression model

It is now possible to return to the LZ78 algorithm. We recall that $M_n$ is the number of segments (or phrases) built on a random text of length $n$. We have:

THEOREM 2. *Let $M_n$ be the number of phrases built on $n$ characters of the text by LZ78, according to the binary memoryless model. Then, the mean and variance of $M_n$ satisfy*

$$\mathrm{E}\, M_n \sim \frac{nh}{\log_2 n}, \qquad \mathrm{Var}\, M_n = \frac{c_2 h^3 n}{\log_2^2 n}.$$

*Furthermore, the distribution of $M_n$ is asymptotically normal,*

$$\frac{M_n - \mathrm{E}\, M_n}{\sqrt{\mathrm{Var}\, M_n}} \xrightarrow{d} \mathcal{N}(0,1).$$

The proof of this result is in [7]. The ideas have been further refined by Louchard and Szpankowski in [8], and this leads to a second order asymptotic analysis, hence a characterization of redundancy of LZ78.

The "inverse" relations alluded to in Section 2 are expressed precisely by the equality,

$$\Pr\{M_n > m\} = \Pr\{L_m \le n\}.$$

This is known as the *renewal equation*. Theorem 1 gives good estimates of the right hand side. Roughly speaking, we have, by Theorem 1, a known dependency between $L$ (path length) and $M$ (size of the dst),

$$L_M \approx \frac{1}{h} M \log_2 M + \sqrt{c_2 M \log_2 M}\, X,$$

where $X$ is a unit Gaussian variate that represents random fluctuations. This can be formally inverted, leading to

$$M \approx \frac{Lh}{\log_2 L} + \sqrt{\frac{c_2 h^3 L}{\log_2^2 L}}\, X.$$

This is exactly what Theorem 2 expresses and the process is made valid by an appeal to the general theory of renewal equations; see [2, Thm. 17.3].

From there, it is possible to solve the redundancy problem of Section 1; see [8] for details. Define the redundancy of LZ78 as

$$\overline{r}_n = \frac{1}{n}\left(\mathrm{E}\{M_n(\log_2 M_n + 1)\} - nh\right).$$

The idea is that there are $M_n$ phrases and each of them costs about $\log_2 M_n$ bits. (The term $+1$, there, is implementation specific but not essential.)

Then moment bounds and inequalities justify a precise version of the approximation

$$\mathrm{E}\{M_n \log_2 M_n\} \approx \mathrm{E}\{M_n\} \log_2 \mathrm{E}\{M_n\}.$$

Also, the speed of convergence to the normal limit in Thm. 1 is seen to be $O(m^{-1/2})$. This gives all the ingredients for a second-order asymptotic analysis of LZ78.

THEOREM 3. *The global redundancy of Lempel-Ziv's LZ78 algorithm is $O(1/\log n)$. More precisely,*

$$\overline{r}_n \sim \frac{1}{\log_2 n}\left(2h - h\gamma - \frac{1}{2}h_2 + h\alpha - h\delta_0(\log_2 n)\right).$$

Note that the corresponding redundancy of LZ77 is known to be

$$\overline{r}_n^* = O(\frac{\log\log n}{\log n}),$$

and this is conjectured to be the right order. Therefore, assuming this conjecture, *the LZ78 algorithm—based on digital search trees and segment boundaries—is less "redundant" and deviates less than LZ77 from the information-theoretic optimum.*

## 5. Related problems

The talk mentions two other types of results:

- an extension to Markovian models, where systems of difference-differential equations need to be considered;
- a modified LZ78 algorithm [8] based on the $b$-digital search tree, a data structure that had been analysed in the average-case by Flajolet and Richmond [4].

One of the major open problems in the area is the precise analysis of redundancy for the LZ77 algorithm. This is harder as the basic digital tree decomposition is no longer available, and "overlaps" in strings must be taken into account. Perhaps the approach of [6] could be useful.

## Bibliography


[1] Aldous (David) and Shields (Paul). – A diffusion limit for a class of randomly-growing binary trees. *Probability Theory and Related Fields*, vol. 79, n° 4, 1988, pp. 509–542.

[2] Billingsley (Patrick). – *Probability and Measure*. – John Wiley & Sons Inc., New York, 1986, second edition, *Wiley Series in Probability and Mathematical Statistics*, xiv+622p.

[3] Flajolet (Philippe), Gourdon (Xavier), and Dumas (Philippe). – Mellin transforms and asymptotics: harmonic sums. *Theoretical Computer Science*, vol. 144, n° 1-2, 1995, pp. 3–58. – Special volume on mathematical analysis of algorithms.

[4] Flajolet (Philippe) and Richmond (Bruce). – Generalized digital trees and their difference-differential equations. *Random Structures and Algorithms*, vol. 3, n° 3, 1992, pp. 305–320.

[5] Jacquet (Philippe) and Régnier (Mireille). – Trie partitioning process: Limiting distributions. In Franchi-Zanetacchi (P.) (editor), *CAAP'86, Lecture Notes in Computer Science*, vol. 214, pp. 196–210. – 1986. Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice, France, March 1986.

[6] Jacquet (Philippe) and Szpankowski (Wojciech). – Autocorrelation on words and its applications: Analysis of suffix trees by string-ruler approach. *Journal of Combinatorial Theory,* Series A, vol. 66, n° 2, 1994, pp. 237–269.

[7] Jacquet (Philippe) and Szpankowski (Wojciech). – Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science*, vol. 144, n° 1–2, 1995, pp. 161–197.

[8] Louchard (Guy) and Szpankowski (Wojciech). – On the average redundancy rate of the Lempel–Ziv algorithm. – Preprint, 1996. 15 pages. Preminary version presented at the *Data Compression Conference*, Snowbird, 1996.

[9] Mahmoud (Hosam M.). – *Evolution of Random Search Trees*. – John Wiley & Sons Inc., New York, 1992, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, xii+324p.

[10] Ziv (Jacob) and Lempel (Abraham). – A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, vol. 23, n° 3, May 1977, pp. 337–343.

[11] Ziv (Jacob) and Lempel (Abraham). – Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, vol. 24, n° 5, September 1978, pp. 530–536.

# Dynamical Systems and Average-Case Analysis of General Tries

*Brigitte Vallée*
GREYC, Université de Caen, France

June 9, 1997

[summary by Julien Clément]

**Abstract**

The three major parameters of a *trie* (sometimes called *digital tree*), number of nodes, path length, and height, are analyzed precisely in a general context where words are emitted by a source associated to a dynamical system. The results can all be stated in terms of two intrinsic characteristics of the source: the entropy and the probability of letter coincidence. These characteristics themselves are linked in a natural way to spectral properties of a Ruelle operator associated to the dynamical system.

## 1. Probabilistic dynamical sources

**1.1. Definitions.** A dynamical source, in the context of information theory, is a mechanism which produces infinite words over an alphabet $\mathcal{M}$. Such a system is defined by four elements: (*i*) an alphabet $\mathcal{M}$ included in $\mathbb{N}$, (*ii*) a quasi partition of $\mathcal{I} = ]0,1[$ with intervals $\mathcal{I}_m$, $m \in \mathcal{M}$, (*iii*) a mapping $\sigma : \mathcal{I} \to \mathcal{M}$ which is constant over each $\mathcal{I}_m$ and equal to $m$ and finally (*iv*) a mapping $T : \mathcal{I} \to \mathcal{I}$ which satisfies two properties: the restriction of $T$ to $\mathcal{I}_m$ is a real analytic bijection from $\mathcal{I}_m$ to $\mathcal{I}$; the mapping $T$ is expansive, i.e. $|T'(x)| > 1$ on $\mathcal{I}$. The words emitted by the source are produced by iterating $T$ and coded thanks to $\sigma$. The word $M(x)$ of $\mathcal{M}^\infty$ (an infinite sequence of symbols), where $x \in \mathcal{I}$, is formed with the symbols

$$M(x) := (\sigma(x), \sigma(T(x)), \sigma(T^2(x)), \ldots, \sigma(T^k(x)), \ldots).$$

Each letter of the alphabet is associated to a distinct branch of $T$ or equivalently to a distinct inverse branch of $T$ denoted by $h_m$. The bijection $h_m : \mathcal{I} \to \mathcal{I}_m$ coincides with the inverse of the restriction of $T$ to $\mathcal{I}_m$.

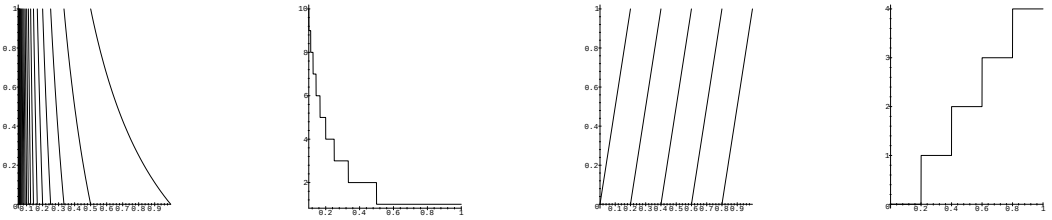


FIGURE 1. Graphical representation of the mappings $T$ and $\sigma$ for a source based on the continued fraction expansion (left) and for a memoryless source based on the 5-ary expansion of numbers (right).

**1.2. Examples.** A lot of probabilistic dynamical sources can be described in such a framework, including all *memoryless sources* where letters of the alphabet can be emitted with probabilities $\{p_i\}$ independently of previous letters. This gives a mapping $T$ where branches are affine. In particular this encompasses the $b$-ary expansion model of numbers. *Markov sources* take into account a finite past for producing words and thus generalize memoryless sources. Finally, in a model where the source is based upon the *continued fraction* expansion of numbers, the alphabet is $\mathbb{N}$ and the probability for a character to be emitted depends on all the previous history.

**1.3. Fundamental intervals, entropy, coincidence probability.** Numbers which share the same prefix expansion $m_1, \ldots, m_k$ form an interval called *fundamental interval* of depth $k$ which is exactly, with the preceding notations,

$$\mathcal{I}_{m_1,\ldots,m_k} = h_{m_1} \circ \cdots \circ h_{m_k}(\mathcal{I}).$$

In a general context, the interval $\mathcal{I}$ is endowed with a continuous density $f$ (so that $F$ denotes the associated distribution). Then the word $M(x)$ is produced according to the expansion process (using $T$ and $\sigma$). In this context, the measure $u_h$ of a fundamental interval $\mathcal{I}_h$ associated to $h = h_{m_1} \circ \cdots \circ h_{m_k}$ is

$$u_h := |F(h(0)) - F(h(1))|,$$

and plays a crucial role since it is the probability that a word begins with a certain prefix. During the analysis, two quantities relative to the source appear naturally. The *entropy* $h(\mathcal{S}, F)$ and the *coincidence probability* $c(\mathcal{S}, F)$ are defined as the limits

$$h(\mathcal{S}, F) := \lim_{k\to\infty} \frac{-1}{k} \sum_{|h|=k} u_h \log u_h, \qquad c(\mathcal{S}, F) := \lim_{k\to\infty} \left[ \sum_{|h|=k} u_h \log u_h \right]^{1/k}.$$

It is interesting to note that these limits exist and are independent of the distribution $F$.

## 2. Tries associated to a general source

Let $\mathcal{M} \in \mathbb{N}$ be a set of elements called digits, and $\mathcal{M}^\infty$ the set of all infinite sequences built over $\mathcal{M}$. For any word produced by the dynamical source $M(x) = (\sigma(x), \sigma(Tx), \sigma(T^2x), \ldots)$ (with $x \in \mathcal{I}$), the head and tail functions are defined by

$$\text{head}(M(x)) = \sigma(x), \quad \text{tail}(M(x)) = M(Tx).$$

Any finite set of infinite words produced by the same source can be written as $M(X) = \{M(x) | x \in X\}$, and one associates to $X$ a trie, Trie$(X)$, defined by the following recursive rules

- (R1) If $X = \{x\}$ has cardinality equal to 1, then Trie$(X)$ consists of a *single leaf node* that contains $M(x)$.
- (R2) If $X$ has cardinality at least 2, then Trie$(X)$ is *an internal node* represented generically by '$o$' to which are attached $\ell$ subtrees, where $\ell = \text{card}\{\sigma(x) | x \in X\}$ is the number of different head symbols in $M(X)$. Let $b_1 < \cdots < b_\ell$ be these head symbols; Trie$(X)$ is defined by

  $$\text{Trie}(X) = \langle o, \text{Trie}(X_1), \ldots, \text{Trie}(X_\ell) \rangle, \quad \text{where} \quad X_j = \{Tx \mid \sigma(x) = b_j, \quad x \in X\}.$$

  The trie Trie$(X_j)$ collects all the suffixes of words that begin with $b_j$.

**2.1. Parameters.** The main parameters of a trie are *size* (number of internal nodes), *height* and *external length path*, which is the sum of all links from the root to each leaf.

The model considers here infinite strings emitted independently by the same dynamical source. Rather than considering a fixed number $n$ of strings, a Poisson model of rate $n$ is used, where the number of strings $N$ is also a random variable which strongly concentrates around $n$. The strong property of independence of this particular model makes the analysis easier and gives access to the expectations of parameters. The expectations of height, size and external path length under a Poisson model of rate $n$ and distribution $F$ over $\mathcal{I}$ are respectively

$$D(n) = \sum_k [1 - \prod_{|h|=k} (1 + n u_h) e^{-n u_h}], \quad S(n) = \sum_h [1 - (1 + n u_h) e^{-n u_h}], \quad P(n) = \sum_h n u_h [1 - e^{-n u_h}].$$

**2.2. Asymptotics.** These quantities are easily recognized as *harmonic sums* of the form $F(x) = \sum_{k \in K} \lambda_k f(\mu_k x)$ (excepted for the height which needs a small calculation step before). The best tool to analyze the asymptotics of such harmonic sums is the *Mellin transform*, which leads to locating poles of the associated *Dirichlet series* $\Lambda(s) = \sum_{k \in K} \lambda_k \mu_k^s$. Here, the key quantity for the analysis of size and path length is the *series of fundamental intervals*,

$$\Lambda(F, s) = \sum_h u_h^s = \sum_h |F(h(0)) - F(h(1))|^s,$$

considered for complex values of $s$. For a general source, it is not always easy (or possible) to locate precisely the singularities. In this case a Tauberian theorem, under some constraints, can be used to extract the asymptotic expansion.

## 3. Generalized Ruelle operators

**3.1. Presentation.** The generalized Ruelle operators are derived from the original Ruelle operators, and involve secants of the inverse branches $H(u, v) := |(h(u) - h(v))/(u - v)|$. The generalized Ruelle operators are defined by

$$\mathbf{G}_s[F](u, v) = \sum_{|h|=1} \tilde{H}(u, v)^s F(h(u), h(s)),$$

where $\tilde{H}$ is the analytic extension of $H$ and $s$ is a complex parameter. Here the sum is taken over branches of depth 1. If we define the secant $L$ of the distribution $F$,

$$L(x, y) = \left| \frac{F(x) - F(y)}{x - y} \right|,$$

then the Dirichlet series can be expressed as

$$\Lambda(F, s) = \sum_h u_h^s = \sum_h |F(h(0)) - F(h(1))|^s = (I - \mathbf{G}_s)^{-1}[L^s](0, 1).$$

**3.2. Singularities of the quasi inverse $(I - \mathbf{G}_s)^{-1}$.** Singularities of $(I - \mathbf{G}_s)^{-1}$ are of special interest because these are also singularities of $\Lambda(F, s)$. These singularities arise for values of $s$ where $\mathbf{G}_s$ has an eigenvalue equal to 1. In particular, there is always a pole at $s = 1$ (easy to prove from the previous form of $\Lambda(F, s)$). One can derive from strong properties of the operator $\mathbf{G}_s$ that there are three different cases, called periodic, quasi-periodic and aperiodic, depending on the precise nature of the eigenvalues of $\mathbf{G}_s$ on the line $\Re(s) = 1$. The operator $\mathbf{G}_s$ has also special properties at $s = 1$ and $s = 2$ since the entropy of the source is $h(\mathcal{S}) = -\lambda'(1)$ (the derivative of the dominant eigenvalue at $s = 1$), while the coincidence probability is $c(\mathcal{S}) = \lambda(2)$.

## 4. Average-case analysis of general tries

**4.1. Analysis of height.** The analysis of height is based on estimates of the individual probabilities $\pi_k(n) = \prod_{|h|=k}(1 + nu_h)e^{-nu_h}$ followed by a Mellin analysis. This leads to the asymptotic expansion

$$D(n) = \frac{2}{|\log c(\mathcal{S})|}\log n + P_F(\log n) + \gamma + A_F + o(1)$$

**4.2. Analysis of size and path length.** The operator $(I - \mathbf{G}_s)^{-1}$ has a simple pole at $s = 1$, and thus gives the main term of the asymptotic expansion. For a general source, a Tauberian theorem can be applied to estimate the contribution of others poles. Finally one has

$$P(n) = \frac{1}{h(\mathcal{S})} n \log n + o(n \log n), \qquad S(n) = \frac{1}{h(\mathcal{S})} n + o(n).$$

## 5. Some important particular cases

**5.1. Bernoulli sources.** The Bernoulli source considers a finite alphabet $\mathcal{M} = \{1, \ldots, r\}$ with probability of emission $\{p_1, \ldots, p_r\}$ (with $p_1 + \cdots + p_r = 1$). In this case the entropy and the coincidence probability have classical expressions

$$H = -\sum_{i=1}^{r} p_i \log p_i, \qquad c = \sum_{i=1}^{r} p_i^2.$$

**5.2. Continued fraction source.** The continued fraction expansion of numbers can be considered as a dynamical source over the infinite alphabet $\mathcal{M} = \mathbb{N}$. The operator of Ruelle is then called the Ruelle-Mayer operator and is defined by

$$\mathcal{G}_s[f](z) = \sum_{m \geq 1} \frac{1}{(m+z)^s} f\left(\frac{1}{m+z}\right).$$

The entropy of the source is linked to the so-called Lévy's constant which plays a central rôle in the the analysis of the Euclidean algorithm. The coincidence probability is a constant which intervenes in two-dimensional generalizations of the Euclidean algorithm. These constants are

$$\lambda'(1) = \frac{\pi^2}{6 \log 2}, \quad \lambda(2) \sim 0.1994.$$

## Bibliography


[1] Brigitte Vallée. – Dynamical systems and average-case analysis of general tries. – Les cahiers du GREYC, 1997.

[2] Flajolet (Philippe) and Vallée (Brigitte). – *Continued Fraction Algorithms, Functional Operators, and Structure Constants.* – Research Report n° 2931, Institut National de Recherche en Informatique et en Automatique, July 1996. 33 pages. Invited lecture at the 7th Fibonacci Conference, Graz, July 1996; to appear in *Theoretical Computer Science.*

[3] Hervé Daudé (Philippe Flajolet) and Vallée (Brigitte). – *An Average-case Analysis of the Gaussian Algorithm for Lattice Reduction.* – Research Report n° 2798, Institut National de Recherche en Informatique et en Automatique, February 1996. To appear in *Combinatorics, Probability and Computing*, 1997.

[4] Jacquet (Philippe) and Szpankowski (Wojciech). – Analysis of digital tries whith Markovian dependency. *IEEE Transactions on Information Theory*, vol. 37, n° 5, September 1991, pp. 1470–1475.

[5] Vallée (Brigitte). – Opérateurs de Ruelle-Mayer généralisés et analyse en moyenne des algorithmes d'Euclide et de Gauss. *Acta Arithmetica*, vol. LXXXI, n° 2, 1997, pp. 101–144.

# Asymptotic Properties of Algorithms for Random Generation of Under-Diagonal Paths

*Guy Louchard*
ULB, Belgium

March 3, 1997

**Abstract**

Using classical properties of random paths and Brownian motion, we obtain the asymptotic distribution of some characteristic quantities of an algorithm proposed by Barcucci et al. to generate under diagonal paths of size $n$. The algorithm is based on a rejection technique.

In terms of random paths, the algorithm is equivalent to analyzing the first meander of length $\geq n$ for a random path reflected at the origin. The meander is defined as a random path starting from the origin and conditioned to remain strictly positive.

We analyze successively the cost $c_n$ defined as the number of steps necessary to obtain a meander of length $n$, the deviation $\delta_n$, which is the height of the meander at its last step and the maximum of the meander.

We obtain asymptotics for various quantities, such as the probability generating function, Laplace transforms of densities, distribution functions, asymptotic densities (Gaussian, exponential, extreme value, Jacobi or Rayleigh), a discrete stationary distribution.

# Algorithms for Variable Length Subnet Address Assignment

*Mike Atallah*
Purdue University

September 23, 1996

**Abstract**

In a computer network that consists of many subnetworks, the binary address of a particular machine consists of two parts: A prefix that contains the address of the subnetwork to which the machine belongs, and a suffix containing the address of that particular machine within its subnetwork. In variable-length subnetwork addressing, the length of the prefix containing the subnetwork's address varies from one subnetwork to another. To avoid ambiguity when decoding addresses, there is a requirement that no subnetwork address be a suffix of another subnetwork address. We give an algorithm for finding a suitable set of subnetwork addresses so as to maximize the total number of addressable machines. We generalize the algorithm to the case where each subnetwork also has a priority associated with it. This is joint work with D. E. Comer.

# Nearest-Neighbour Search in High Dimension and Molecular Clustering

*Frédéric Cazals*
Algorithms project, INRIA Rocquencourt

June 30, 1997

[summary by Frédéric Cazals]

## 1. Introduction and prerequisites

**1.1. Problem statement.** Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$, the nearest-neighbour (NN) and $k$ nearest neighbours ($k$-NN) problems can be stated as follows: pre-process $P$ in order to return as fast as possible the nearest or $k$ nearest neighbour(s) of an arbitrary point $q$ according to any Euclidian metric $d(p,q) = (\sum_{i=1}^{d}(p_i - q_i)^2)^{1/2}$. A weakened version of the NN problem consists in returning a point $p'$ which $\varepsilon$-approximates the NN $p$ of $q$ in the sense $d(p',q)d(p,q) \leq 1+\varepsilon$ for any $\varepsilon > 0$. If one denotes $p_{i_1}, \ldots, p_{i_n}$ the points of $P$ sorted by increasing distance to $q$, an equivalent formulation for the $k$-NN problem consists in returning a subset $S = \{s_1, \ldots, s_k\}$ with $d(q, s_j) \leq (1+\varepsilon)d(q, p_{i_j})$ for $j = 1, \ldots, k$.

The naive algorithm to compute the NN of a point $q$ consists in checking all the points of $P$ and returning the closest, which has complexity $O(dn)$. On the other hand, the most sophisticated algorithms known until recently had complexities in $O(exp(d)\log n)$ with $exp(d)$ a function growing at least as quickly as $2^d$—see e.g., [1]. So that whenever $d \geq \log n$ nothing better than the brute force method was known!

Kleinberg's break-through [4] has been to get around the exponential difficulty by an heavy use of random sampling aiming at "comparing" the points of $P$ through their projections on random lines passing through the origin rather than decomposing the $d$-dimensional space containing them. The first result is an algorithm returning an approximation of the $k$-NN in a deterministic way but with an exponential time/space pre-processing. The second algorithm returns an approximation of the $NN$ in a randomised way but with a polynomial pre-processing only. This talk presents these two algorithms and discusses their potential use to a clustering problem arising in chemistry—see section 4.

**1.2. Prerequisites.**

1.2.1. *A geometric lemma.* The core idea of Kleinberg's method lies in the following property:

LEMMA 1. *Let $x$ and $y$ be two vectors of $\mathbb{R}^d$ such that $\| y \| / \| x \| \geq 1+\gamma$ with $\gamma \leq 1/2$. Then, if $v$ is a random vector on the unit sphere $S^{d-1}$ we have* $\Pr[| x \cdot v | \geq | y \cdot v |] \leq 1/2 - \gamma/3$.

Intuitively, short vectors "should not defeat too often" longer ones when comparing their projection on a random line determined by a vector on $S^{d-1}$. In order to compare two vectors from their projections, the key point is therefore to use a large enough set of lines to capture the probabilistic property contained in the above theorem.

1.2.2. *Empirical measures and Vapnik-Chervonenkis bounds.* Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $S \subset \mathcal{F}$ a set of events, and $X_1, X_2, \ldots, X_n$ $n$ random variables following the law $\mu$. If one calls the empirical measure of an event $s$ the fraction of $X_i$'s falling into $s$, the quantity

$$\Delta_S^n = \sup_{s \in S} \left| \frac{1_s(X_1) + \cdots + 1_s(X_n)}{n} - \mu(s) \right|$$

measures the maximum difference over the class $S$ between the empirical measure and the probability. It is a random variable and Vapnik-Chervonenkis's contribution [5] has been to elucidate the conditions under which it converges in probability to zero, that is the conditions under which $\lim_{n \to \infty} \Pr[\Delta_S^n > \varepsilon] = 0$ for any $\varepsilon$. To sketch this contribution, let a range-space be a couple $(\mathcal{P}, \mathcal{R}) = \big((\Omega, \mathcal{F}, \mu), \mathcal{R} \subset \mathcal{F}\big)$. We shall say that a set $A$ of finite cardilality is shattered by $\mathcal{R}$ if $\forall a \in 2^A \quad \exists r \in \mathcal{R}$ such that $a = r \cap A$. The dimension of Vapnik-Chervonenkis of $(\mathcal{P}, \mathcal{R})$ is the cardinality of the biggest $A \subset \Omega$ shattered by $\mathcal{R}$.

DEFINITION 1. A $\gamma$-sample for $(\mathcal{P}, \mathcal{R})$ is a finite set $A \subset \Omega$ such that $|\mu(r) - \mid r \cup A \mid / \mid A \mid| \leq \gamma$, $\forall r \in \mathcal{R}$.

THEOREM 1 ([5]). *For a range space of dimension $k$, a random sample of size*

$$l \geq \frac{16}{\gamma^2}(k \log \frac{16k}{\gamma^2} + \log \frac{4}{\delta})$$

*is a $\gamma$-sample with a probability at least $1 - \delta$.*

1.2.3. *Exceptional and $\rho$-distinguishing sets.* As pointed out above, we are interested in comparing points with respect to their projections on vectors from $S^{d-1}$. For two vectors $x$ and $y$ with $\| y \| / \| x \| \geq 1 + \gamma$ we call their exceptional set

$$W_{x,y} = \{v \in S^{d-1} \text{ such that } \mid x \cdot v \mid \geq \mid y \cdot v \mid\}.$$

And a random set of vectors $V$ from $S^{d-1}$ is called $\rho$-distinguishing if

$$\forall W_{x,y}, \quad \mu(W_{x,y}) < \rho \Longrightarrow \mid V \cap W_{x,y} \mid / \mid V \mid < 1/2.$$

More prosaically, a set $V$ is $\rho$-distinguishing if a majority of its points do not fall into some exceptional set of size smaller than $\rho$.

1.2.4. *Hyper-planes arrangements.* An arrangement of $n$ hyper-planes in $\mathbb{R}^d$ is said to be in general position if any $d$ hyper-planes have a unique point in common, and any $d+1$ hyper-planes do not share a point. Given a hyper-plane $h$ and a point $p$, $p$ is either above, on, or below $h$, which is called its position. A face on an arrangement is the set of points having the same position with respect to all the hyper-planes. The dimension of a face is its affine dimension. It is known from [2] that

THEOREM 2. *The number of $d$-faces of an arrangement of $n$ hyper-planes in general position is* $f_d(n) = \sum_{i=0}^{d} \binom{n}{i}$.

1.2.5. *Digraphs.* A complete digraph $G$ on $n$ vertices $1, 2, \ldots, n$ is a directed graph which contains for any pair of vertices $\{i, j\}$ either the edge $(i, j)$ or $(i, j)$. An apex of $G$ is a vertex with a directed path of length at most two to any vertex. At least, an apex ordering of $G$ is an ordering $i_1, \ldots, i_n$ of its vertices such that $i_k$ is an apex for the sub-digraph $G[i_k, i_{k+1}, \ldots, i_n]$. The following is straightforward:

THEOREM 3. *Every $n$-node complete digraph has an apex ordering computable in $O(n^2)$.*

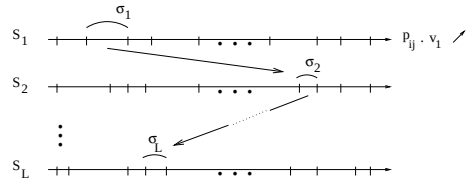



FIGURE 1. Traces

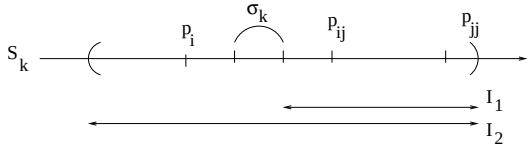



FIGURE 2. $\sigma$-domination

**1.3. First results.** The following theorems can be proved:

LEMMA 2. *Let $\mu$ be the uniform measure on $S^{d-1}$. The dimension of the range-space*

$$((S^{d-1}, \mathcal{F}, \mu), \{W_{x,y} \mid \mu(W_{x,y}) \leq \rho\})$$

*is less than $d' = 8(d+1)\log(4d+4)$.*

LEMMA 3. *There exists $c_0$ such that a random sample of $S^{d-1}$ of size $f(\delta,\gamma)$ is a $\gamma/2$-sample for the range-space $((S^{d-1}, \mathcal{F}, \mu), \{W_{x,y} \mid \mu(W_{x,y}) \leq \rho\})$ with $f(\delta,\gamma) = \frac{c_0}{\gamma^2}\left(d' \log \frac{d'}{\gamma^2} + \log \frac{1}{\delta}\right) = \theta(d \log^2 d)$.*

COROLLARY 1. *A set $V$ of $f(\gamma,\delta)$ vectors from $S^{d-1}$ is $(1/2-\gamma)$-distinguishing with a probability at least $1-\delta$.*

## 2. First algorithm

**2.1. Construction of the data structure.** This algorithm returns an approximation of the $k$-NN of a point $q$. To build the data structure from which it does so, we first draw uniformly at random a set $V$ of $L = f(\frac{\varepsilon}{3}, \delta) = \theta(d \log^2 d)$ vectors from $S^{d-1}$. Then for each vector $v_l \in V$, the following is done: 1. compute $v_l \cdot p_{ij}$ with $p_{ij} = (p_i + p_j)/2$, $1 \leq i, j \leq n$; 2. sort the $p_{ij}$ according to the values of $v_l \cdot p_{ij}$ and denote $S_l$ the list obtained. The list of lists $S_1, \ldots, S_L$ is denoted $\Sigma$.

For each such list, a pair of consecutive entries is called a primitive interval, and a sequence of primitive intervals is called a trace—see Figure 1. The maximum number of traces is upper-bounded by $(n^2)^L = n^{O(d \log^2 d)}$. But a trace is realizable if

$$\exists q \in \mathbb{R}^d, \forall k = 1, \ldots, L, v_k \cdot p^{(k)}_{i_1 i_2} < v_k \cdot q < v_k \cdot p^{(k)}_{i_3 i_4}.$$

So that realizable traces are defined with respect to the $Ln^2$ hyper-planes $v_k \cdot (p^{(k)}_{i_1 i_2} - x)$. And from theorem 2, the number of such traces is $\sum_{i=0}^{d} \binom{Ln^2}{i} = O(n \log d)^{2d}$.

DEFINITION 2. For a realizable trace $\sigma = \sigma_1 \cdots \sigma_k \cdots \sigma_L$, $p_i$ is said to $\sigma$-dominate $p_j$ in $S_k$ if $p_{ij}$ lies in the interval $(\sigma_k, \ldots, p_{jj})$.

For each realizable trace $\sigma$, the construction is as follows: (1.) build a complete digraph $G_\sigma$ on $\{1, 2, \ldots, n\}$ with the edge $(i,j)$ if $p_i$ $\sigma$-dominates $p_j$ in half of the lists of $\Sigma$; (2.) build an apex ordering $(\sigma, G_\sigma)$ of the nodes of $G_\sigma$.

The idea behind the domination definition is depicted on Figure 2: if $p_{ij}$ falls in the desired interval denoted $I_1$, then $p_i \in I_2$ and we have $\mid v_k \cdot (p_i - q) \mid < \mid v_k \cdot (p_j - q) \mid$.

**2.2. Algorithm.** To process a query associated to a point $q$: 1. compute $\sigma(q) = \sigma_1(q) \cdots \sigma_L(q)$ with $\sigma_k(q)$ the primitive interval from $S_l$ containing $v_l \cdot q$; 2. retrieve the apex ordering associated to $\sigma(q)$ and return the $k$ first entries.

This algorithm therefore returns an $\varepsilon$-approximation of the $k$ nearest neighbours of a point $q$ in a deterministic fashion, so that the answer is guaranteed to be correct if the random sample $V$ is

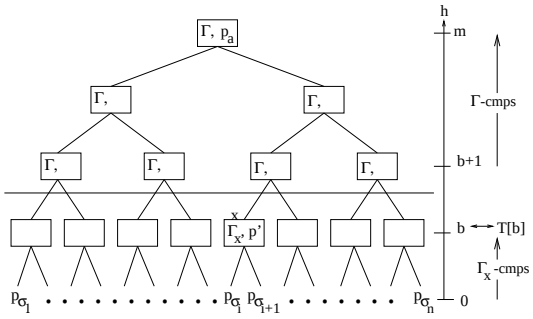



FIGURE 3. Tournament tree

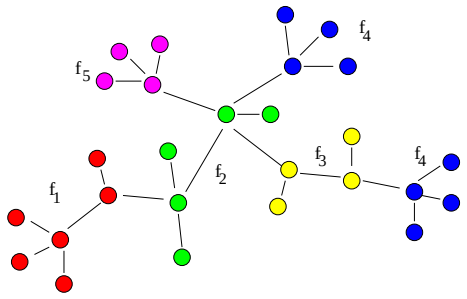



FIGURE 4. Molecule and fragments

actually $(1/2 - \varepsilon/3)$-distinguishing—see §1.2.3 and Corollary 1. Roughly speaking, the correctness of the algorithm comes from the fact that if a non-desired point $p_1$ has been returned instead of a desired point $p_2$, then $p_1$ dominated $p_2$ in more than half of the lists of $\Sigma$, and the sample $V$ was not distinguishing enough with respect to the exceptional set $W_{q-p_1,q-p_2}$.

The pre-processing requires $O\left(Ln^2(n\log d)^{(2d)}\right)$ time and $O\left(n(n\log d)^{(2d)}\right)$ space. The cost of a query is $O(k + L(d + \log n)) = O(k + (d\log^2 d)(d + \log n))$.

## 3. Second algorithm

As opposed to the first algorithm, the second one does not try to compute a partition of the requests' space and proceeds in two steps. The first one consists in drawing a random sample from $P$ of the appropriate size and returning the closest point to $q$. The second one compares iteratively $q$ to pairs of points of $P$ in a tournament way depicted on Figure 3. Assuming that $n = 2^m$, the overview of this second stage is the following:

1. a random sample $V$ of $\Theta(d\log^2 n(\log^2 d + \log d\log\log n))$ vectors is drawn uniformly on $S^{d-1}$, and a multi-set $\Gamma_v$ of $V$ is assigned to each internal node $v$ of the binary tree whose leaves are a permutation of the points of $P$. For a query point $q$, two points $p_i$ and $p_j$ of $P$, and a multi-set $\Gamma_v$, $p_i$ is said to dominate $p_j$ if $\mid v_k \cdot (p_i - q) \mid < \mid v_k \cdot (p_j - q) \mid$ holds for a majority of vectors $v_k$ in $\Gamma_v$. Otherwise $p_j$ dominates $p_i$;
2. each internal node $v$ of the tree is assigned its dominating child for the multi-set $\Gamma_v$.

The point eventually returned is the best candidate from the two points returned by the two steps. It can be shown that an $\varepsilon$-approximation is returned with a probability greater than $1 - \delta$ in $O(n + d\log^3 n)$ time with a space requirement of $n \mid V \mid$.

## 4. Application to molecular clustering

Suppose we are given a set of $d$ molecular fragments, say $F = \{f_1, \ldots, f_d\}$ and a set $M = \{m_1, \ldots, m_{N_m}\}$ of $N_m$ molecules, each described by a set of fragments of $F$—see Figure 4. We shall represent a molecule $m_i$ by a sequence of $d$ boolean values $m_i = b_{i,1}b_{i,2}\cdots b_{i,d}$ with $b_{i,j} = 1$ if $m_i$ contains $f_j$ and $b_{i,j} = 0$ otherwise. This formulation fails to report multiple occurrences of a given fragment in a molecule, but it has the nice property that a molecule is represented by a point on the hyper-cube $H^d = \{0,1\}^d$. Given two molecules, we call their *similarity* the number of common fragments that is the quantity $\mathrm{sim}(m_i, m_j) = \sum_{k=1}^{d} b_{i,k} \cdot b_{j,k}$.

Given the set $M$ we are interested in the following problem: find a partition of $M$ into subsets of neighbours or clusters. To do so, one way to proceed see [3] consists in first building a Minimum Spanning Tree on the input data set, second removing those "too long" edges from the MST,

and third computing the connected components we are left with. The key point lies in the MST computation, and it is shown in [6] that

THEOREM 4. *A MST on $n$ points in dimension $d$ can be found in $O(2^d n^{2-1/2^{d+1}}(\log n)^{1-1/2^{d+1}})$ time.*

Unfortunately for our concern where $d$ can range from 500 to 2000, the $2^d$ constant is prohibitive. Kleinberg's algorithms customised to the hyper-cube setting could make Yao's algorithm interesting in practice.

## Bibliography


[1] Arya (S.), Mount (D. M.), Netanyahu (N. S.), Silverman (R.), and Wu (A. Y.). – An optimal algorithm for approximate nearest neighbour searching. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. pp. 573–582. – New York, 1994.

[2] Edelsbrunner (Herbert). – *Algorithms in combinatorial geometry*. – Springer-Verlag, Berlin, 1987, *EATCS Monographs on Theoretical Computer Science*, vol. 10, xvi+423p.

[3] Jain (Anil K.) and Dubes (Richard C.). – *Algorithms for clustering data*. – Prentice-Hall Inc., Englewood Cliffs, NJ, 1988, *Prentice-Hall Advanced Reference Series*, xiv+320p.

[4] Kleinberg (J.). – Two algorithms for nearest-neighbour search in high dimension. In *ACM STOC*. – El Paso, Texas, USA, 1997.

[5] Vapnik (N.) and Chervonenkis (A.). – On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, vol. 16, n° 2, 1971, pp. 264–280.

[6] Yao (Andrew Chi Chih). – On constructing minimum spanning trees in $k$-dimensional spaces and related problems. *SIAM Journal on Computing*, vol. 11, n° 4, 1982, pp. 721–736.

# Part 4

# Probabilistic Methods

# Wiener-Hopf Factorization: Probabilistic Methods

*Philippe Robert*
Inria Rocquencourt

March 17, 1997

[summary by Jean-François Dantzer]

## 1. Introduction

We consider a discrete random walk on $\mathbb{Z}$, defined by

$$S_0 = 0 \qquad \text{and} \qquad S_n = \sum_{i=1}^{n} X_i, \quad n > 0,$$

where $(X_i)_{i\geq 1}$ is an independent identically distributed (i.i.d.) sequence of random variables. We define two hitting times $\nu_+$ and $\nu_-$,

$$\nu_+ = \inf\{k > 0 / S_k > 0\}, \qquad \nu_- = \inf\{k > 0 / S_k \leq 0\},$$

with the convention $\inf(\emptyset) = +\infty$.

We also define $M$ and $L$ two variables indicating respectively the maximum of the random walk and the hit moments at which it is attained

$$M = \sup_{n\geq 0}\{S_n\}, \qquad L = \inf_{n\geq 0}\{S_n = M\}.$$

This talk presents the classical probabilistic methods to derive the join distributions of $(\nu_+, S_{\nu_+})$, $(\nu_-, S_{\nu_-})$ and $(M, L)$.

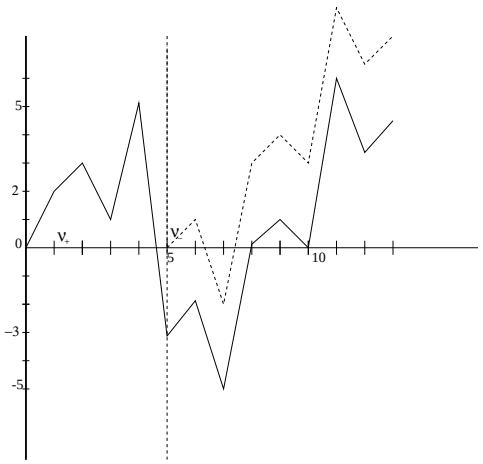



FIGURE 1. Random walks $(S_n)_{n\geq 0}$ and $(S_{n+\nu_-})_{n\geq 0}$ (in dotted line)

Applications of these results have been found in biology [3, 4] or in queueing theory [2]. For more details on this subject, see [1, 5].

## 2. Distribution of $(\nu_+, S_{\nu_+})$ and $(\nu_-, S_{\nu_-})$

The distributions of the pairs $(\nu_+, S_{\nu_+})$ and $(\nu_-, S_{\nu_-})$ will be expressed through their generating functions, i.e., $E(u^{\nu_+} z^{S_{\nu_+}})$ and $E(u^{\nu_-} z^{S_{\nu_-}})$.

We consider three variables

$$\Psi_+(u,z) = \frac{1}{1 - E(u^{\nu_+} z^{S_{\nu_+}})} \qquad \text{on} \qquad \{|u| < 1, |z| \leq 1\},$$

$$\Psi_-(u,z) = \frac{1}{1 - E(u^{\nu_-} z^{S_{\nu_-}})} \qquad \text{on} \qquad \{|u| < 1, |z| \geq 1\},$$

$$\Psi(u,z) = \sum_{n \geq 0} E(u^n z^{S_n}) \qquad \text{on} \qquad \{|u| < 1, |z| = 1\}.$$

The main result, the factorization of Wiener-Hopf described in the following proposition, gives an analytic characterization of $\Psi_+$ and $\Psi_-$.

PROPOSITION 1. *$\Psi$ can be uniquely decomposed on $\{|z| = 1\}$ as:*

$$\Psi(u,z) = \Psi_+(u,z)\Psi_-(u,z),$$

*where $\Psi_+$ and $\Psi_-$ have the following properties:*

- *$\Psi_+$ is analytic on $\{|z| < 1\}$;*
- *$\Psi_+$ and $1/\Psi_+$ are continuous and bounded on $\{|z| \leq 1\}$;*
- *$\Psi_+(u,0) = 1$;*

*and*

- *$\Psi_-$ is analytic on $\{|z| > 1\}$;*
- *$\Psi_-$ and $1/\Psi_-$ are continuous and bounded on $\{|z| \geq 1\}$.*

PROOF. The proof is based on the following arguments:

- The function $\Psi$ can be expressed with the distribution of $X_1$, the sequence $(X_i)_{n \geq 1}$ is independent identically distributed, then $E(Z^{S_n}) = E(Z^{\sum_{i=1}^n X_i}) = E(Z^{X_1})^n$, thus

$$\Psi(u,z) = \sum_{n \geq 0} u^n E(z^{X_1})^n = \frac{1}{1 - uE(z^{X_1})};$$

- $(S_n)_{n \geq 0}$ and $(S_{n+\nu_-})_{n \geq 0}$ have same distribution, $(S_{n+\nu_-})_{n \geq 0}$ is the random walk beginning at the time $\nu^-$;
- The independence between the pair $(\nu_-, S_{\nu_-})$ and the sequence $(S_{n+\nu_-})_{n \geq 0}$;
- the same properties are also valid for $(S_{n+\nu_+})_{n \geq 0}$.

□

## 3. Examples

The factorization is easy when $\Psi$ has a finite number of poles and zeros. We consider two such examples.

**3.1. Random walk ± 1.** We suppose $X_i = 1$ with probability $p$ and $X_i = -1$ with probability $(1-p)$. In that case,

$$\Psi(u,z) = \frac{z}{-upz^2 + z - u(1-p)}$$

$\Psi$ has only two poles

$$\alpha_1(u) = \frac{1 - \sqrt{1-4u^2p(1-p)}}{2up}, \qquad \alpha_2(u) = \frac{1 + \sqrt{1-4u^2p(1-p)}}{2up},$$

$$\text{with} \qquad 0 \le \alpha_1(u) \le 1 \le \alpha_2(u).$$

The decomposition gives:

$$\Psi_+(u,z) = \frac{\alpha_2(u)}{\alpha_2(u) - z} \qquad \text{and} \qquad \Psi_-(u,z) = \frac{z}{\alpha_2(u)up(z - \alpha_1(u))}.$$

Here, we obtain the generating functions:

$$E(u^{\nu_+} z^{S_{\nu_+}}) = \frac{z}{\alpha_2(u)}, \qquad E(u^{\nu_-} z^{S_{\nu_-}}) = (1 - \alpha_2(u)up) + \frac{u(1-p)}{z}.$$

**3.2. Random walk left bounded.** We suppose $\Pr(X_i < -1) = 0$.

$$\Psi(u,z) = \frac{1}{1 - uE(z^X)} = \frac{z}{z - uE(z^{X+1})}.$$

In that case, the factorization is easy because by Rouché's theorem the function $z \mapsto z - uE(z^{X+1})$ has one only root which belongs to $\{|z| < 1\}$, which we denote by $\alpha(u)$. One proves that $\alpha(u) \in [0,1]$ and the decomposition of $\Psi$ is the following:

$$\Psi_+(u,z) = \frac{z - \alpha(u)}{z - uE(z^{X+1})} \frac{u\Pr(X=-1)}{\alpha(u)}, \qquad \Psi_-(u,z) = \frac{z}{z - \alpha(u)} \frac{\alpha(u)}{\Pr(X=-1)},$$

and the generating functions are:

$$E(u^{\nu_-} z^{S_{\nu_-}}) = 1 - \frac{u\Pr(X=-1)}{\alpha(u)} + \frac{u\Pr(X=-1)}{z},$$

$$E(u^{\nu_+} z^{S_{\nu_+}}) = 1 - \frac{\alpha(u)}{u\Pr(X=-1)} + \frac{z - uE(z^{X+1})}{z - \alpha(u)}.$$

## 4. Distribution of the Maximum and its first Hitting time $(M, L)$

The distribution of the pair $(M, L)$ is expressed through its generating function $E(x^L z^M)$.

PROPOSITION 2.

$$E(x^L z^M) = \lim_{u \to 1} \frac{\Psi_+(ux, z)}{\Psi_+(u,1)}.$$

PROOF. We define the variables $M_n$ and $L_n$, as

$$M_n = \max_{0 \le k \le n} S_k, \qquad L_n = \inf\{k / S_k = M_n\},$$

and the function $H$ on $\{|u| < 1, |x| < 1, |z| < 1\}$ as

$$H(u,x,z) = E(\sum_{n \ge 0} u^n x^{L_n} z^{M_n}).$$

Using the same arguments as in proposition 1, it can be proved that:

$$H(u,x,z) = \frac{\Psi_+(ux,z)}{\Psi_+(u,1)(1-u)}.$$

We conclude applying

$$\lim_{u\to 1}(1-u)H(u,x,z) = \lim_{u\to 1}(1-u)E(\sum_{n\geq 0} u^n x^{L_n} z^{M_n}) = \lim_{n\to+\infty} E(x^{L_n} z^{M_n}) = E(x^L z^M).$$

□

For the case of the random walk of subsection 3.1, it gives for $p < \frac{1}{2}$:

$$E(x^L z^M) = \frac{\alpha_2(u)}{\alpha_2(u) - z}\frac{1-2p}{1-p},$$

then

$$\Pr(L < +\infty, M < +\infty) = 1,$$

and the distribution of $M$ is geometric with parameter $\frac{p}{1-p}$, for $p > \frac{1}{2}$:

$$E(x^L z^M) = 0 \qquad \text{and} \qquad \Pr(L = M = +\infty) = 1.$$

## Bibliography


[1] Feller (William). – *An introduction to probability theory and its applications.* – John Wiley & Sons, New York, 1971, 2nd edition, vol. II.

[2] Iglehart (Donald L.). – Extreme values in the $GI/G/1$ queue. *Annals of Mathematical Statistics*, vol. 43, n° 2, 1972, pp. 627–635.

[3] Karlin (Samuel) and Altschul (Stephen F.). – Methods for assesing the statistical significance of molecular sequences features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the USA*, vol. 87, 1990, pp. 2264–2268.

[4] Karlin (Samuel) and Dembo (Amir). – Strong limit theorems of empirical functionals for large exedances of partial sums of i.i.d. variables. *Annals of Probability*, vol. 19, n° 4, 1991, pp. 1737–1755.

[5] Spitzer (F.). – *Principles of Random Walk.* – Van Nostrand, 1964.

# Wiener-Hopf Factorization and Maximal Scores in Biological Sequences

*Pierre Nicodème*
INRIA-Rocquencourt

February 10, 1997

[summary by Philippe Robert]

## 1. Introduction

In this talk we study a matching problem for two sequences $S = (s_1, \ldots, s_n)$, $T = (t_1, \ldots, t_p)$ where $s_1, \ldots, s_n, t_1, \ldots, t_p$ are elements of some alphabet $A$. This mathematical model is used in the analysis of some biological sequences. To any couple $(x, y) \in A \times A$, one associates a score $H(x, y) \in \mathbb{R}$ which is negative if the two letters do not agree or positive if their affinity is significant. A local matching of length $l$ of these sequences is given by a sequence $((s_{i_1}, t_{j_1}), (s_{i_1}, t_{j_1}), \ldots, (s_{i_1}, t_{j_1}))$ with $1 \leq i_1 < i_2 < \cdots < i_l \leq n$ and $1 \leq j_1 < \cdots < j_l \leq p$. The score of this matching is then defined as

$$\sum_{k=1}^{l} H(s_{i_k}, t_{j_k}).$$

The main problem considered in this talk is to estimate the maximal score among all the possible matchings of these sequences.

A probabilistic setting is used to give estimates of this optimal score. The letters are assumed to be drawn independently from the alphabet. This hypothesis leads to a formulation of the problem in terms of random walk. The optimal score $M(n)$ for two sequences of size $n$ can be represented as

$$M(n) = \sup_{0 \leq j \leq k \leq n} (S_k - S_j), \tag{1}$$

where $S_n = \sum_{i=1}^{n} X_i$ ($\sum_1^0 = 0$); The variables $(X_i)$ are assumed to be independent and identically distributed. The sequence $(S_n)$ is the random walk starting from 0 associated to the distribution of $X_1$. Clearly the sequence $(M(n))$ is non decreasing with $n$, and as we will see, it converges to infinity as $n \to +\infty$. Our goal is to find an asymptotic estimate of $M(n)$ for $n$ large. We prove that if $E(X_1) < 0$, and some other technical conditions, there exists some constant $\alpha$ such that the renormalized sequence $M(n) - \alpha \log(n)$ converges in distribution.

## 2. The relation with a reflected random walk

For $n \geq 1$ we denote by

$$W_n = \sup_{0 \leq k \leq n} (S_n - S_k),$$

then $M(n)$ can be expressed as

$$M(n) = \sup_{0 \leq k \leq n} W_k. \tag{2}$$

It is easy to see that the sequence $(W_n)$ satisfies the following relation

$$W_{n+1} = (W_n + X_{n+1})^+, \qquad n \geq 0 \tag{3}$$

where $x^+ = \max(x, 0)$. Now define

$$\nu_- = \inf\{n > 0 / S_n \leq 0\},$$

which is the first time the random walk visits the negative axis. The law of large numbers gives that almost surely $\lim_{n\to+\infty} \frac{1}{n}\sum_1^n X_i = E(X_1)$, and because $E(X_1) < 0$, we have $\lim_{n\to+\infty} \sum_1^n X_i = -\infty$, thus the quantity $\nu_-$ is always finite.

By induction, using (3), one can check that $W(n) = S_n$ for $n < \nu_-$. Furthermore, we have $W_{\nu_-} = (S_{\nu_-})^+ = 0$, by definition of $\nu_-$. It is easy to prove that starting from $t = \nu_-$, the sequence $W_n$ performs another similar excursion above 0 independently of the previous excursion, and so on. The sequence $(W_n)$ is the reflected random walk at 0.

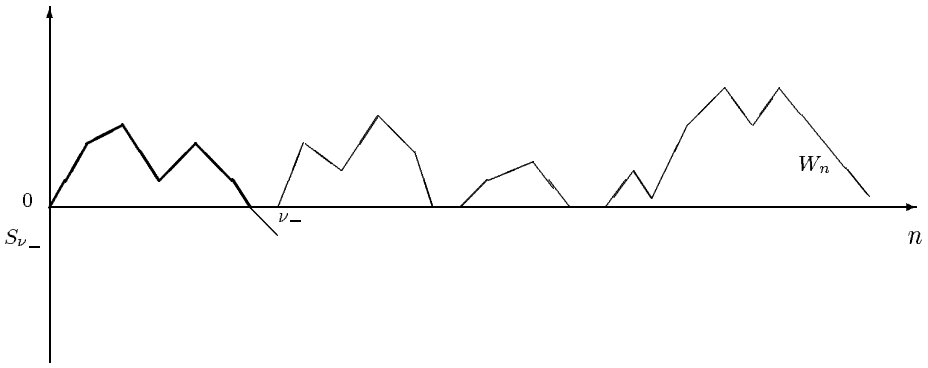



FIGURE 1. The path of $n \to W_n$

*The method of resolution.* To estimate $M(n)$, the maximum of $k \to W_k$ on the interval $\{0, \ldots, n\}$, we proceed as follows

1. Estimate the distribution of $M_{exc}$, the maximum of the random walk during an excursion above 0.
2. Count the number of excursions of $k \to W_k$ above 0 up to time $n$.
3. $M(n) = \max\{M_{exc_i}\}$ where the maximum is taken on all excursions $exc_i$ before time $n$. The excursions being independent, the same is true for the $(M_{exc_i})$. Estimating the maximum of independent random variables is easy.

Technically, the main tool used to prove convergences is the renewal theorem. The explicit calculation of constants involved in the limiting distribution requires the Wiener-Hopf factorization associated to this random walk. We give a formulation of these results in the next section.

## 3. The probabilistic tools

**3.1. The renewal theorem.** We consider a sequence of non negative i.i.d. integrable random variables $(Y_i)$ and denote by $T_n = \sum_1^n Y_i$, the non decreasing random walk associated to the distribution of $Y_1$. For $t \in \mathbb{R}_+$, let $N_t$ be the number of $T_k$ between 0 and $t$. Thus, $T_{N_t+1} - t$ is the length of the interval between $t$ and the first $T_k$ after $t$.

PROPOSITION 1. *Almost surely*

$$\lim_{t\to+\infty} \frac{N_t}{t} = \frac{1}{E(Y_1)},$$

*where $E(Y_1)$ denotes the expected value of $Y_1$.*

As we can remark at that point the above proposition will solve the second point of our program above. In this case the beginnings of the excursions will define the renewal process.

The main result in renewal theory concerns the solution of the so-called renewal equation. If $f$ is some function, the function $Z$ is the solution of the renewal equation associated to $f$ if, for all $x \geq 0$,

$$Z(x) = f(x) + \int_0^x Z(x-y)P(X_1 \in dy). \tag{4}$$

The main theorem is the following

THEOREM 1. *If $f$ is Riemann integrable, there is a unique solution $Z_f$ of (4) and $Z_f$ satisfies*

$$\lim_{x \to +\infty} Z_f(x) = 1/E(X_1) \int_0^{+\infty} f(u)du.$$

This analytical formulation of the renewal theorem can be seen as a consequence of a probabilistic result: the variable $T_{N_t+1} - t$ converges in distribution as $t \to +\infty$.

**3.2. The Wiener-Hopf factorization.** This technique concerns the calculation of the distribution of the hitting times of the positive, negative axis by a random walk and the position of the random walk at these times. We have already seen $\nu_-$, we define its positive counterpart $\nu_+$,

$$\nu_+ = \inf\{n/S_n > 0\}.$$

THEOREM 2. *For $u \in \mathbb{C}$, such that $|u| < 1$, there exist $\phi_+(u,.)$, $\phi_-(u,.)$ such that*

1.

$$\frac{1}{1 - uE(e^{-\xi X})} = \phi_+(u,\xi)\phi_-(u,\xi), \qquad \Re(\xi) = 0. \tag{5}$$

2. *The function $\phi_+(u,.)$ [resp. $\phi_-(u,.)$] is analytic on $\{\Re(\xi) > 0\}$ [resp. $\{\Re(\xi) < 0\}$], continuous, bounded away from 0 and $\infty$ on $\{\Re(\xi) \geq 0\}$ [resp. $\{\Re(\xi) \leq 0\}$]. Moreover*

$$\lim_{\Re(\xi) \to +\infty} \phi_+(u,\xi) = 1.$$

*Such a decomposition is unique.*

The following corollary is the probabilistic interpretation of the above theorem.

COROLLARY 1. *The functions of the Wiener-Hopf factorization can be expressed as*

$$\phi_+(u,\xi) = \frac{1}{1 - E(u^{\nu_+} e^{-\xi S_{\nu_+}})}, \qquad |u| < 1, \quad \Re(\xi) \geq 0,$$

$$\phi_-(u,\xi) = \frac{1}{1 - E(u^{\nu_-} e^{-\xi S_{\nu_-}})} \qquad |u| < 1, \quad \Re(\xi) \leq 0.$$

Thus, if we are able to decompose the function $1/(1 - uE(e^{-\xi X}))$, the joint distributions of $(\nu_+, S_{\nu_+})$ and $(\nu_-, S_{\nu_-})$ are known through their Fourier-Laplace transforms, $E(u^{\nu_+} e^{-\xi S_{\nu_+}})$, $E(u^{\nu_-} e^{-\xi S_{\nu_-}})$.

## 4. The main results

Using theorem 1, one can prove the proposition about the tail distribution of the maximum of the random walk during an excursion.

PROPOSITION 2. *If the following conditions are satisfied,*

- $E(X_1) < 0$, $P(X_1 > 0) > 0$ *and* $X_1$ *is non arithmetic;*
- *There exists* $\theta > 0$ *such that* $E(e^{\theta X_1}) < +\infty$*;*
- $E(|X_1| e^{\gamma X_1}) < +\infty$, *where* $\gamma$ *is the positive solution of* $E(e^{\gamma X_1}) = 1$*;*

*then*

$$\lim_{x \to +\infty} e^{\gamma x} P\left(M_{exc} \geq x\right) = C_{exc} = \frac{P(\nu_+ = +\infty)(1 - E(e^{\gamma S_{\nu_-}}))}{\gamma E(X_1 e^{\gamma X_1}) E(\nu_+ e^{\gamma S_{\nu_+}} 1_{\{\nu_+ < +\infty\}})}.$$

Notice that to make the constant $C_{exc}$ explicit, one has to know some functionals of $\nu_+$, $\nu_-$. This is the place where the Wiener-Hopf decomposition is useful.

At that point cases 1 and 2 of our program are solved. For point 3 it remains to integrate these results. This gives our final theorem.

THEOREM 3. *Under the hypotheses of the proposition 2, there exist two constant* $K, \lambda$ *such that*

$$\lim_{n \to +\infty} P(M(n) - \frac{\log(n)}{\lambda} \leq x) = e^{-Ke^{-\lambda x}}.$$

### Bibliography


[1] Asmussen (Søren). – *Applied Probability and Queues.* – John Wiley & Sons, Chichester, 1987, *Wiley Series in Probability and Mathematical Statistics.*

[2] Feller (William). – *An introduction to probability theory and its applications.* – John Wiley & Sons, New York, 1971, 2nd edition, vol. II.

[3] Iglehart (Donald L.). – Extreme values in the $GI/G/1$ queue. *Annals of Mathematical Statistics*, vol. 43, n° 2, 1972, pp. 627–635.

[4] Karlin (Samuel) and Altschul (Stephen F.). – Methods for assesing the statistical significance of molecular sequences features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the USA*, vol. 87, 1990, pp. 2264–2268.

[5] Karlin (Samuel) and Dembo (Amir). – Strong limit theorems of empirical functionals for large exedances of partial sums of i.i.d. variables. *Annals of Probability*, vol. 19, n° 4, 1991, pp. 1737–1755.

# The Philosophers' Process on Graphs

*Bernard Ycart*
INRIA-Rocquencourt

November 18, 1997

[summary by Philippe Robert]

## 1. Introduction

The talk is concerned with a stochastic model based on the dining philosophers' problem which is called the philosophers process. The philosophers are on the $N$ vertices of some graph $G$, they have two states 0 or 1. The state of the process is given by the vector $\eta = (\eta(x))$ where $\eta(x) \in \{0,1\}$ is the state of node $x$. The dynamics of this process forbid the states where two vertices with value 1 are neighbours.
The stochastic process is described as follows:

- If a node has the value 1, then at rate 1 it becomes 0;
- If a node and all its neighbours have the value 0, then at rate $\lambda$, its value becomes 1.

This can be translated in a Markov process context; If $\eta_t \in \{0,1\}^N$ is the state of the process at time $t$, then its generator $Q = (q(\eta,\zeta))$ is given by

1. $q(\eta, \eta - \delta_x) = 1$ if $\eta(x) = 1$, where $\delta_x(y) = 1$ if $x = y$ and 0 otherwise;
2. $q(\eta, \eta + \delta_x) = \lambda$ if $\eta(y) = 0$ for all $y \in N(x)$, where $N(x)$ is the set of the neighbours of $x$;
3. $q(\eta,\zeta) = 0$ otherwise if $\eta \neq \zeta$.

Because of the constraint on the 1's, only a subset $S$ of $\{0,1\}^N$ contains the states which are admissible. It is well known that this process is reversible, that is at equilibrium the process from time 0 to time $T$ looks the same in distribution as in the reversed time scale from $T$ to 0. Its equilibrium measure $\pi$ is also known, $\pi(\eta) = [\lambda^{\sum_x \eta(x)}]/Z_G$, with $\eta \in S$. $Z_G$ is a normalizing constant, so that $\pi(S) = 1$, also known as the partition function. The constant $Z_G$ is a polynomial in $\lambda$, its degree $I$ corresponds to the states with a maximal number of nodes with value 1. The coefficient of $\lambda^I$ in $Z_G$ is the number of states with $I$ nodes with value 1. Hence, in some sense, finding the explicit value of $Z$ can be presented as a combinatorial problem. The main results of this talk concern the explicit representations that can be obtained for the equilibrium measure and in particular for the partition function.

## 2. Markov random fields

We begin with some general definitions concerning Markov random fields. These objects are probability measures on the state space which can be expressed in terms of local specifications instead of a global description such as the one we have encountered for $\pi$. It turns out that the measure $\pi$ as we shall see can also be expressed in that setting. This result gives a simple recursive sequence to express the partition function $Z_G$.

DEFINITION 1. (i). A probability measure $\mu$ on $\{0,1\}^N$ is a *Markov Random Field* if for all $U, V \subset \{1,\ldots,N\}$ such that $V \cup N(V) \subset U$, and $\eta$, $\mu(\eta^U) = \mu(\eta^{U-V})\mu(\eta^V/\eta^{N(V)})$, where $\eta^U$ is the set of coordinates of $\eta$ whose index is in $U$ and $N(V)$ is the set of the neighbours of $V$. (ii). A probability measure $\mu$ on $\{0,1\}^N$ is a *clique system* if for all $U \subset \{1,\ldots,N\}$, $\eta \in S$,

$$\mu(\eta^U) = \frac{1}{Z}\prod_C Q_C(\eta^C),$$

where the product is on the subsets $C$ of the set of indices such that $C$ is a complete subset of $U$ or $C$ is in the complementary of $U$.

Obviously (ii) is stronger than (i). Roughly speaking, the definition (i) expresses that the state on $V$ can be determined by the neighbours of $V$, in this sense $\mu$ has local specifications.

PROPOSITION 1. *The measure $\pi$ defined above is a clique system.*

Using the definition of Markov Random Fields, the following proposition holds. It gives a recursive procedure to calculate the partition function.

PROPOSITION 2. *If $C \subset \{1,\ldots,N\}$ is a complete subgraph of $G$, then $Z_G = Z_{G-x} + \lambda Z_{G-\bar{x}}$, $Z_G = Z_{G-C} + \lambda \sum_{x\in C} Z_{G-\bar{x}}$, $\mu_G(\eta(x)=0) = Z_{G-x}/Z_G$, $\mu_G(\eta(x)=1) = Z_{G-\bar{x}}/Z_G$.*

## 3. Applications to some graphs

If $G_1$ and $G_2$ are two graphs, the Cartesian product of these graphs $G_1 \times G_2$ is the graph for which the set of vertices is the usual Cartesian product of the set of vertices. Two nodes are neighbours, if one their coordinates is the same and the two other ones are neighbours in the corresponding initial graph. The graph in line [resp. on the torus] with $m$ nodes is denoted as $L_m$ [resp. $C_m$].

PROPOSITION 3. *If $K_n$ is the clique with $n$ vertices, then the partition function of the graphs $K_n \times C_m$, $K_n \times L_m$ can be expressed explicitly.*

Some recursive procedures can also be defined in the case of hypercubes and trees.

## 4. A phase transition phenomenon

Up to now, the state space of our processes were finite. The Markov processes we considered were irreducible, i.e. with positive probability one can reach one state from any other state. As a consequence, a unique equilibrium measure existed. If the graph has an infinite number of nodes, this is not the case anymore. It is possible to have more than one equilibrium measure, and even an infinite number. In this case the model is said to have a phase transition. Although it occurs with infinite graphs, it also has consequences for finite graphs. Roughly speaking, if there are at least two invariant measures for the infinite graph, this implies that for a finite graph converging to this graph, the equilibrium measure will be a combination of at least two attractors for the state process. Here, we are interested by the Cartesian product $K_n \times T^d_\infty$, where $T^d_\infty$ is the infinite regular tree of degree $d$.

PROPOSITION 4. *If $d \geq n \geq 2$ and $\lambda > (d+1)^d/[d^d(d-n+1)]$, then the associated Markov process has more than one equilibrium measure.*

## Bibliography


[1] Louth (G. M.). – *Stochastic Networks: Complexity, Dependence and Routing.* – PhD thesis, Cambridge University, December 1990.

[2] Ycart (B.). – The philosophers' process: an ergodic reversible nearest particle system. *Annals of Applied Probability*, vol. 3, n° 2, 1993, pp. 356–363.

# The Load Transfer Model

*Bernard Ycart*
INRIA-Rocquencourt

November 18, 1996

[summary by Philippe Robert]

## 1. Introduction

This talk considers a problem of load sharing. A set of $N$ processors are the vertices of some graph $G$ and each of them has a buffer of finite capacity $K$. A flow of requests (which may be thought of as tasks to be processed) arrives at each vertex. In the case where the vertices do not cooperate in some way, this model is simply a set of $N$ independent finite capacity buffers. In order to optimize the use of total capacity of the network, it is desirable to design a policy which transfers a fraction of the load of highly loaded nodes to under-used processors.

These problems are generic in the sense that they occur very frequently in many mathematical models as soon as finite capacity and possibility of cooperation is assumed. A classical example is the telephone network: If a call from $A$ wants to reach $B$ and all the links between $A$ and $B$ are used, it is possible to use the alternative routing strategy: a node $C$ is chosen at random, a link between $A$ and $C$ an a link between $C$ and $B$ are reserved if possible. If this succeeds, the call between $A$ and $B$ is finally accepted. Notice however that this algorithm has a cost: two links are required instead of one [1]. Another example, again for routing, is Valiant's algorithm [2]. It is often quite easy to guess some reasonable load sharing algorithm, however it is more difficult to estimate its impact on the performances of the system. Moreover, in some cases, load sharing is worst than no sharing at all [1, 4]! In this talk, our basic assumptions are: *(i).* The requests arrive at each node according to a Poisson process with parameter $\lambda$; *(ii).* Each processor works at unit speed the requests which are assumed to be exponentially distributed with parameter 1; *(iii).* A request arriving at a node which is already full (i.e. with $K$ requests) is rejected.

To estimate the performances of a load sharing algorithm, one can use the following parameters: the rejection probability; the average use of each processor, or equivalently the average free space in the buffer; the mean response time.

## 2. The load sharing algorithm

We describe the state of the network as a vector $(\eta(x))$ on the set of vertices, $\eta(x)$ is the number of requests at node $x$. The algorithm considered here is defined as follows: If a request arrives at some node $x$, then

1. if $\eta(x) = K$, the request is lost;
2. If $\eta(x) < K$ and $\eta(x) \leq \inf\{\eta(y)/y \in N(x)\}$, where $N(x)$ denotes the set of neighbours of $x$, then the request is assigned to the node $x$ and $\eta(x) \to \eta(x) + 1$;
3. If $K > \eta(x) > \inf\{\eta(y)/y \in N(x)\}$, then the request is sent to a node $Y$, chosen at random in the set $\{y \in N(x)/\eta(y) < \eta(x)\}$. At node $Y$, the request follows the same procedure and 3.

is applied until case 2. is satisfied. Hence the request will finally be assigned to a processor which is a local minimum in terms of the number of requests in the buffer.

A departure from the node $x$ has the same consequences: a request chosen at random among the nodes with a higher load level is transferred to the node $x$ and so on.

The probabilistic assumptions make the state process $(\eta_t(x))$ at time $t$, a Markov process. The transition rates can be defined as follows: Let $\alpha(x,\eta)$ be the rate of the requests *through* node $x$ when the state of the process is $\eta$, then $\alpha(x,\eta)$ is defined recursively:

1. If $\eta(x) \geq \sup\{\eta(y)/y \in N(x)\}$, then $\alpha(x,\eta) = \lambda$ if $\eta(x) < K$, $\alpha(x,\eta) = 0$ otherwise.
2. $\eta(x) < \sup\{\eta(y)/y \in N(x)\}$, then $\alpha(x,\eta) = \lambda + \sum_{y \in N(x), \eta(y) > \eta(x)} \alpha(y,\eta)/\underline{n}(y,\eta)$, where $\underline{n}(y,\eta) = |\{z/\eta(z) < \eta(y)\}|$.

These transitions are quite straightforward according to the description we gave. The term $\frac{1}{\underline{n}(y,\eta)}$ is a consequence of the fact that if $y$ is not a local minimum, then the request is sent at random to some neighbour with a lower load.

The parameter $\alpha(y,\eta)$ is *not* the real rate of requests in the buffer of $y$. This rate is 0 if $y$ is not a local minimum and $\alpha(y,\eta)$ otherwise. The departure rates are defined in a similar way, using the local minima of $\eta$.

## 3. Results

The Markov process we defined above has a local interaction, i.e. any event concerns only a bounded neighbourhood. The local interaction comes from the fact that a local minimum is at most at distance $K$ from any node. This is a special case of interacting particle systems [3]. One of the first non trivial problems for this class of models is the existence of such a process. Given a generator, that is, the transitions, does there exist a semi-group, i.e. a Markov process, with this generator? Standard techniques show that this is the case. The second problem is the asymptotic behaviour: does there exist an equilibrium measure, a unique one? and if not what are the extreme points of the convex set of equilibrium measures? In the case where the graph is the lattice $\mathbb{Z}^d$, the result is the following

THEOREM 1. *There is a unique equilibrium measure and the state of the network converges exponentially fast to that equilibrium measure.*

The proof of this theorem is based on a coupling argument: if $\eta$, $\zeta$ denote two possible states for the network such that $\eta \leq \zeta$, i.e. $\eta(x) \leq \zeta(x)$ for all the nodes $x$, then it is possible to construct two processes $\eta_t$, $\zeta_t$ with initial state $\eta$, $\zeta$ such that $\eta_t \leq \zeta_t$ for all $t \geq 0$. The proof uses this coupling to prove that all the processes stick to the stationary process after some time $T$ with an exponential moment.

**3.1. The case of the complete graph.** When the nodes are all connected, it is then easy to see that the total number of requests $X(t)$ in the system is a birth and death process with birth rate $n\lambda$ if $X(t) \leq (K-1)N$ and $(KN-j)\lambda$ if $(K-1)N < X(t) \leq KN-1$. The death rates are given by $X(t)$ if $X(t) \leq N$, and $N$ otherwise. Asymptotic results are easily derived when $N \to +\infty$.

## Bibliography


[1] Kelly (F. P.). – Loss networks. *Annals of Applied Probability*, vol. 1, n° 3, 1991, pp. 319–378.

[2] Leighton (F. T.). – *Parallel algorithms and architectures: array-trees-hypercubes.* – Morgan Kaufman, 1992.

[3] Liggett (T. M.). – *Interacting Particle Systems.* – Springer Verlag, New York, 1985, *Grundlehren der mathematischen Wissenschaften.*

[4] Malyshev (V. A.) and Robert (P.). – Phase transition in a loss load sharing model. *Annals of Applied Probability*, vol. 4, n° 4, 1994, pp. 1161–1176.

# Probability and Number Theory: Some Examples of Connections

*Jean-Marc Deshouillers*
Mathématiques Stochastiques, Université Bordeaux 2

January 13, 1997

[summary by Alain Plagne]

**Abstract**

We illustrate some connections between probability theory and number theory. Examples are taken from multiplicative number theory (probabilistic behaviour of the function $\omega$), additive number theory (Sidon problem, where arithmetic results are proved by using probability theory; partitions; and Erdős-Rényi type models for sums of powers, where situations on which nothing is known, are modelled) and probability (upper bound for the concentration of the sum of independent and identically distributed integer random variables, where probabilistic results are obtained by using methods from additive number theory).

## 1. Introduction

One can distinguish at least four types of connections between probability theory and additive number theory.

(i) When probability leads to models for the integers;
(ii) When probability techniques permit to prove number theory results;
(iii) When number theory questions lead to probability questions;
(iv) When number theory methods permit to prove probability results.

Part 2 is devoted to multiplicative number theory. It provides examples for points (i) and (ii). The third part, which is the heart of this talk, is concerned with additive number theory and illustrated with Sidon's problem (point (ii)), partitions (points (i) and (ii)) and sums of $s$-th powers (points (i) and (iii)). In the fourth part, which deals with concentration functions, point (iv) is illustrated.

## 2. Multiplicative Number Theory

We begin with a famous result due to Hardy and Ramanujan [11] concerning the function $\omega$, which counts the number of divisors of an integer

$$\omega(n) = |\{p \text{ such that } p|n\}|.$$

Here and in the sequel, $p$ always denotes a prime number.

THEOREM 1. *Let $\Psi(x) \to \infty$. Then the set of integers such that*

$$|\omega(n) - \log\log n| \leq \Psi(n)\sqrt{\log\log n} \tag{1}$$

*has density one.*

This result means that

$$\lim_{N\to+\infty} \frac{|\{1 \le n \le N : (1) \text{ holds}\}|}{N} = 1.$$

The original proof is quite technical. We quote here an efficient way to get the result which is due to Turán [17].

$$N^{-1}\sum_{n\le N}\omega(n) = N^{-1}\sum_{n\le N}\sum_{p|n} 1 = N^{-1}\sum_{p\le N}\sum_{\substack{n\le N\\ n=0 \bmod p}} 1$$
$$= N^{-1}\sum_{p\le N}(N/p + O(1)) = \sum_{p\le N} 1/p + O(1) = \log\log N + O(1).$$

See for example [16] for the last equality. In the same way, we get

$$N^{-1}\sum_{n\le N}\omega(n)^2 = (\log\log N)^2 + O(\log\log N),$$

thus the number of exceptions to (1) up to $N$ is

$$|\{1 \le n \le N : |\omega(n) - \log\log n| \ge \Psi(n)\sqrt{\log\log n}\}| \le \sum_{n\le N} \frac{(\omega(n) - \log\log n)^2}{\Psi(n)^2 \log\log n},$$

which is $O(N/\Psi(N)^2)$. This proves the result.

Now let us show a probabilistic approach. Let $(X_p)$ be a family of independent random variables defined by

$$X_p = \begin{cases} 1, & \text{with probability } 1/p, \\ 0, & \text{with probability } 1 - 1/p. \end{cases}$$

The mathematical expectation of $\omega = \sum_{p\le N} X_p$ is

$$E(\omega) = \sum_{p\le N} E(X_p) = \sum_{p\le N} 1/p = \log\log N + o(\log\log N),$$

and the variance of $\omega$ is

$$V(\omega) = \sum_{p\le N} V(X_p) = \sum_{p\le N} (1 - 1/p)/p = \log\log N + o(\log\log N).$$

Chebyshev's inequality yields

$$\Pr\left\{|\omega - E(\omega)| \ge \Psi\sqrt{V(\omega)}\right\} \le 1/\Psi^2,$$

that is the result.

This result can be interpreted, with a probabilistic point of view, as a weak law of large numbers. A much more precise result, that is a central limit theorem, has been proved by Erdős and Kac in 1939 [4, 5].

THEOREM 2. *For $u$ a real, one has*

$$\frac{1}{N}|\{1 \le n \le N : \omega(n) - \log\log n \le u\sqrt{\log\log n}\}| \underset{N\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-t^2/2}dt. \tag{2}$$

A natural question is now: can this result be extended to other (strongly) additive functions (that is such that $f(n) = \sum_{p|n} f(p)$)? More precisely, can one compare (and how far?)

$$\frac{1}{N}|\{1 \le n \le N : f(n) \le u\}| \qquad \text{with} \qquad \Pr\left\{\sum_{p \le N} X_p \le u\right\},$$

where the $X_p$'s are independent Bernoulli random variables such that $\Pr\{X_p = f(p)\} = 1/p$ and $\Pr\{X_p = 0\} = 1 - 1/p$. By sophisticated sieve arguments, it can be shown that

$$\frac{1}{N}|\{1 \le n \le N : \sum_{p|n, p \le r} f(p) \le u\}| - \Pr\{\sum_{p \le r} X_p \le u\} = O\left[x^{-1/15} + \exp\left(-\frac{1}{8}\frac{\log x}{\log r}\log\frac{\log x}{\log r}\right)\right],$$

which is $o(1)$ when $r = x^{\epsilon(x)}$ tends to $+\infty$ but $\epsilon(x) \to 0$. This model is called the "Kubilius model".

## 3. Additive Number Theory

**3.1. Sidon problem.** In 1932, Sidon raised the following question: is it possible to find a sequence $\mathcal{A}$ of non-negative integers such that

(i) For every positive integer $n$, there exist $a, b$ in $\mathcal{A}$ such that $n = a + b$;
(ii) $|\{a, n - a \in \mathcal{A}\}| = o(n^\epsilon)$ for any $\epsilon > 0$.

In other words: is it possible to build a "thin" basis of order 2? In 1954, Erdős [3] answered positively the question, by proving a more precise result.

THEOREM 3. *There exists $\mathcal{A}$ and $0 < c_1 < c_2$ such that*

$$c_1 \log n < |\{a, n - a \in \mathcal{A}\}| < c_2 \log n, \tag{3}$$

*for $n \ge 2$.*

The proof is probabilistic and particularly short but absolutely not constructive: let $(\Omega, T, P)$ be a probabilistic space and $X_2, X_3, \ldots$ be independent Bernoulli random variables such that $\Pr\{X_n = 1\} = c\sqrt{(\log n)/n}$. For $\omega \in \Omega$, let $\mathcal{A}(\omega) = \{0, 1\} \cup \{n : X_n(\omega) = 1\}$. Then for almost all $\omega \in \Omega$, $\mathcal{A}(\omega)$ satisfies (3). A very detailed proof of this result is given in the book by Halberstam and Roth [9, chap. 3].

**3.2. Restricted partition function.** This example underlies the probabilistic interpretation of the powerful Ramanujan-Hardy-Littlewood circle method [18].

For a given $N$, let $q(N)$ be the number of ways to write

$$N = n_1 + n_2 + \cdots + n_r,$$

with $0 < n_1 < n_2 < \cdots < n_r (\le N)$. Another way to say this is: let

$$\mathcal{E}_N = \{0, 1\} \times \cdots \times \{0, N\},$$

then

$$q(N) = 2^N \frac{|\{x \in \mathcal{E}_N, \sum x_n = N\}|}{|\{x \in \mathcal{E}_N\}|}.$$

Let $X_1, \ldots, X_N$ be independent random variables such that $X_n$ takes the values 0 and $n$ with probability $1/2$. We have

$$q(N) = 2^N \Pr\{X_1 + \cdots + X_N = N\}.$$

Denote $E_N$ and $V_N$ the expectation and the variance of $X_1 + \cdots + X_N$. If we assume that a local limit theorem holds, then we can write

$$q(N) = 2^N \frac{1}{\sqrt{2\pi V_N}} \left( \exp\left( -\frac{(N - E_N)^2}{2V_N} \right) + o(1) \right).$$

An easy computation shows that $E_N \sim N^2/4$, $V_N \sim N^3/12$ and that the exponential term is $o(1)$ and so that $q(N) = o(2^N N^{-3/2})$ which is not interesting. This is due to the fact that $N$ is too far from $E_N$.

This approach is far from being perfect. The reason is quite clear: in the problem of partitions for an integer $N$, the integers $1, 2, \ldots, N$ do not have the same importance: the small values take part much more frequently in a decomposition of $N$ than the large ones. Thus we have to refine the model by weighting the different integers, with a smaller weight for large integers.

Suppose now the $X_i$'s are independent random variables taking only the values 0 and $n$ with $\Pr\{X_n = n\} = p_n$. Then

$$q(N) = \frac{\sum_{x \in \mathcal{E}_n, x_1 + \cdots + x_N = N} \Pr\{(X_1, \ldots, X_N) = x\}}{\prod_{n, x_n = n} p_n \prod_{n, x_n = 0} (1 - p_n)}.$$

If we take $p_n = \exp(-\sigma n)/(1 + \exp(-\sigma n))$, for some $\sigma$ such that $E(X_1 + \cdots + X_N) = N$, that is to say

$$\sigma = \frac{\pi}{2\sqrt{3N}}(1 - 1/8N + O(N^{-2})),$$

we can prove (this is naturally the heart of the matter) that

$$q(N) \sim \frac{1}{4(3N)^{1/4}} \exp(\pi N^{1/2}/\sqrt{3})$$

(see [10, 13]). This argument has been recently developed in [8] in a more general context.

Proof goes as usual, by defining the characteristic function (i.e., the Fourier transform of the image measure)

$$\phi_n(t) = (1 - p_n) + p_n \exp(2\pi i n t).$$

Then

$$\Pr\{X_1 + \cdots + X_N = N\} = \int_{\mathbb{R}/\mathbb{Z}} \left( \prod_{n \le N} \phi_n(t) \right) \exp(-2\pi i N t)\, dt.$$

The main term (corresponding to major arc) comes from a neighbourhood of 0 on the torus $\mathbb{R}/\mathbb{Z}$. There "remains" (it is not easy!) to find an upper bound for $|\prod_{n \le N} \phi_n(t)|$ outside this neighbourhood, that is on the minor arc.

**3.3. Probabilistic models for sums of powers.** It is well known that sums of two squares have zero density. But for $s \ge 3$, nothing is known about (lower) density of the set of sums of $s$ integral $s$-th powers, like sums of 3 cubes and of 4 biquadrates.

There are two conjectures. In 1968, Barrucand [1] conjectured that for $s = 3$ and 4 the answer is NO. But, in 1986, Hooley [12] conjectured that the answer is YES for every $s \ge 3$.

In order to guess something in this hard problem, Erdős and Rényi [6], in 1960, considered "pseudo $s$-th powers", i.e., random sequences $\mathcal{A}^{(s)}$ defined as in the answer to Sidon's question, and suggested that the number of representation $r_s(N)$ of an integer $N$ as a sum of $s$ elements from $\mathcal{A}^{(s)}$ should follow (almost surely) a Poisson law. Unfortunately, their proofs contained a gap because of the difficulty of dealing with the quasi-independence of the sets involved. In 1965, Halberstam and Roth [9] overcame the difficulty when $s = 2$, by combinatorial arguments. In 1995, Landreau [14]

proved a correlation inequality, having its own probabilistic interest, which leads to the expected result.

THEOREM 4. *Let $E_1, \dots, E_N$ be independent events, and $A_1, \dots, A_T$ be such that each $A_t$ is an intersection of some of the $E_n$'s. Then*

$$0 \leq \Pr(\cap \bar{A}_t) - \prod \Pr(\bar{A}_t) \leq \sum_{1 \leq t < t' \leq T} (\Pr(A_t \cap A_{t'}) - \Pr(A_t)\Pr(A_{t'})).$$

## 4. Probability

We illustrate here the possibility of applying number theory ideas to probability theory with the following recent theorem taken from [2], of which we sketch the proof.

THEOREM 5. *Let $\frac{\log 4}{\log 3} < \sigma \leq 2$, $n \in \mathbb{N}^*$ and $\epsilon > 0$, $X_1, \dots, X_n$ be i.i.d. integral valued random variables such that*

$$\max_{q \leq 2} \max_{s \bmod q} \sum_{l = s \bmod q} \Pr\{X_1 = l\} \leq 1 - \epsilon \tag{4}$$

*and for all $L \geq 2$*

$$Q(X_1, L) \leq 1 - L^{-\sigma}, \tag{5}$$

*where $Q(Y, x)$ denotes $\sup_{h \in \mathbb{R}} \Pr\{h < Y \leq h + x\}$. Then there exists a constant $c = c(\sigma, \epsilon, Q(X_1, 1))$ such that*

$$Q(X_1 + \cdots + X_n; 1) \leq c n^{-1/\sigma}.$$

The proof, in the same manner as above, uses the characteristic function $\phi$ of $X_1$. If $S_n = X_1 + \cdots + X_n$, we have

$$\Pr\{S_n = k\} = \int_0^1 \phi(t)^n \exp(-2\pi i k t) dt,$$

so that $Q(S_n, 1) \leq \int_0^1 |\phi(t)|^n dt$.

We are now reduced to study the large values of $\phi$. We first use a lemma which has been introduced in [15] (and which follows from Bochner's theorem).

LEMMA 1. *Let $E_\theta = \{t \in \mathbb{R}/\mathbb{Z}, |\phi(t)| \geq \cos\theta\}$, where $\mathbb{R}/\mathbb{Z}$ is once again the torus, we have for $\theta_1, \theta_2 \geq 0$ and $\theta_1 + \theta_2 \leq \pi/2$,*

$$E_{\theta_1} + E_{\theta_2} \subset E_{\theta_1 + \theta_2}.$$

Then we need a result by Freiman [7] on structure of set addition (it has been extended to the torus in [15]).

THEOREM 6. *Let $\mathcal{A}$ be a finite subset of $\mathbb{Z}$ such that*

$$|\mathcal{A} + \mathcal{A}| \leq 2|\mathcal{A}| - 1 + b,$$

*with $b \leq |\mathcal{A}| - 3$, then there exists an arithmetic progression $\mathcal{L}$ with $|\mathcal{A}| + b$ elements that contains $\mathcal{A}$.*

This enables us to show that either the set of the arguments for which $\phi$ is large has a small measure, or it has a structure (it is located close to the vertices of a regular polygon), which is not possible in view of (4) and (5).

## Bibliography


[1] Barrucand (Pierre). – Sur la distribution empirique des sommes de trois cubes ou de quatre bicarrés. *Comptes-Rendus de l'Académie des Sciences*, vol. 267, 1968, pp. 409–411.

[2] Deshouillers (J.-M.), Freiman (G. A.), and Yudin (A. A.). – *On Bounds for the Concentration Function, 1.* – Prépublication n° M/95/37, IHES, 1995.

[3] Erdős (P.). – On a problem of Sidon in additive number theory. *Acta Scientiarum Mathematicarum Szegediensis*, vol. 15, 1954, pp. 255–259.

[4] Erdős (P.) and Kac (M.). – On the Gaussian law of errors in the theory of additive functions. *Proceedings of the National Academy of Sciences of the USA*, vol. 25, 1939, pp. 206–207.

[5] Erdős (P.) and Kac (M.). – The Gaussian law of errors in the theory of additive number theoretic functions. *American Journal of Mathematics*, vol. 62, 1940, pp. 738–742.

[6] Erdős (P.) and Rényi (A.). – Additive properties of random sequences of positive integers. *Acta Arithmetica*, vol. 6, 1960, pp. 83–110.

[7] Freĭman (G. A.). – *Foundations of a structural theory of set addition.* – American Mathematical Society, Providence, R. I., 1973, *Translations of Mathematical Monographs*, vol. 37. Translated from the Russian.

[8] Freiman (Gregory A.) and Pitman (Jane). – Partitions into distinct large parts. *Journal of the Australian Mathematical Society*, vol. 57, n° 3, 1994, pp. 386–416.

[9] Halberstam (Heini) and Roth (Klaus Friedreich). – *Sequences.* – Springer-Verlag, New York-Berlin, 1983, 2nd edition.

[10] Hardy (G. H.) and Ramanujan (S.). – Asymptotic formulæ in combinatory analysis. *Proceedings of the London Mathematical Society*, vol. 16, 1917, pp. 75–115.

[11] Hardy (G. H.) and Ramanujan (S.). – The normal number of prime factors of a number $n$. *Quarterly Journal of Mathematics*, vol. 48, 1917, pp. 76–92.

[12] Hooley (Christopher). – On some topics connected with Waring's problem. *Journal für die Reine und Angewandte Mathematik*, vol. 369, 1986, pp. 110–153.

[13] Hua (Loo-Keng). – On the number of partitions of a number into unequal parts. *Transactions of the American Mathematical Society*, vol. 51, 1942, pp. 194–201.

[14] Landreau (Bernard). – Étude probabiliste des sommes de $s$ puissances $s$-ièmes. *Compositio Mathematica*, vol. 99, n° 1, 1995, pp. 1–31.

[15] Moskvin (D. A.), Freĭman (G. A.), and Judin (A. A.). – Inverse problems of additive number theory and local limit theorem for lattice random variables. In *Number-theoretic studies in the Markov spectrum and in the structural theory of set addition*, pp. 148–162. – Kalinin. Gos. Univ., Moscow, 1973.

[16] Tenenbaum (Gérald). – *Introduction to analytic and probabilistic number theory.* – Cambridge University Press, Cambridge, 1995, *Cambridge Studies in Advanced Mathematics*, vol. 46. Translated from the second French edition (1995).

[17] Turán (P.). – On a theorem of Hardy and Ramanujan. *Journal of the London Mathematical Society*, vol. 9, 1934, pp. 274–276.

[18] Vaughan (R. C.). – *The Hardy-Littlewood method.* – Cambridge University Press, Cambridge, 1997, 2nd edition, *Cambridge Tracts in Mathematics*, vol. 125.

# Sums of Cubes: Algorithmic and Numerical Aspects

*François Hennecart*
A2X, Université Bordeaux 1

January 13, 1997

[summary by Alain Plagne]

**Abstract**

Here are presented results of joint work by J.-M. Deshouillers, F. Hennecart and B. Landreau on sums of powers (and especially of three and four cubes): do they have a positive density? is their behaviour that of the probabilistic model? Moreover, they exhibit a candidate for being the largest integer which is not sum of four cubes, namely 7 373 170 279 850.

## 1. Sums of cubes

In 1770, Waring wrote that every integer is the sum of 4 squares, 9 cubes, 19 biquadrates and so on, meaning that for each integer $k$, there exists a constant $g(k)$ such that every integer $N$ is the sum of at most $g(k)$ $k$-th powers. It was not until 1909, that Hilbert [11] proved it, by a difficult argument.

We say that an integer is $C_k$ if it is the sum of at most $k$ cubes. In 1912, Kempner and Wierferich proved that every integer is $C_9$, that is sum of at most 9 cubes. In 1939, Dickson [7] proved that, except 23 and 239, every integer is $C_8$. Later, Linnik [15] (and later Watson [21] and Mac Curley [16]), proved that every sufficiently large integer is $C_7$. Papers by Bohman and Fröberg [2] and Romani [17] suggest that there are only 15 integers $C_8$ and not $C_7$ (the largest one being 454), 121 that are $C_7$ and not $C_6$ (the largest one being 8042), and 3922 that are $C_6$ and not $C_5$ (the largest one being 1290740).

The circle method, introduced and developed by Hardy, Littlewood and Ramanujan [10] yields an asymptotic formula for the number of solutions to some Diophantine equations. It gives, for large enough $s$,

$$\mathcal{R}_s(N) = |\{0 \leq x_1, \ldots, x_s \leq N, N = x_1^3 + \cdots + x_s^3\}| \sim \mathcal{S}_s(N)\frac{\Gamma(4/3)^s}{\Gamma(s/3)}N^{s/3-1} \tag{1}$$

when $N$ tends to $+\infty$. The factor $\mathcal{S}_s(N)$ is commonly called the singular series

$$\mathcal{S}_s(N) = \sum_{q=1}^{\infty} \sum_{\substack{a \bmod q \\ (a,q)=1}} q^{-s} S(a,q)^s e_q(-aN),$$

where $e_q(u) = \exp(2\pi i u/q)$ and

$$S(a,q) = \sum_{m=1}^{q} e_q(am^k).$$

This singular series reflects the arithmetic properties of sums of cubes and usually does not imply difficulty because (when it is convergent) it can be written as an Eulerian product (that is a

product over primes). In 1985, Vaughan [19] proved that (1) holds true for $s \geq 8$ and two years later, showed [20] the lower bound

$$\mathcal{R}_7(N) \gg \mathcal{S}_7(N) N^{4/3}.$$

The usual conjecture is that (1) is true as soon as $s \geq 4$. In this direction, Davenport [4] proved, in 1939, that

$$E(N) = |\{n \leq N, \text{ such that } n \text{ is not } C_4\}| \ll_\epsilon N^{\frac{29}{30}+\epsilon},$$

which implies that almost every integer is $C_4$. Recently the exponent has been reduced to 37/43 [3].

We denote by $\mathcal{R}'_3(n)$ the number of solutions of $x^3 + y^3 + z^3 = n$, with $0 \leq x \leq y \leq z$. It is clear that the number $f_3(N)$ of integers $n \leq N$ which are sums of three positive cubes (that is such that $\mathcal{R}'_3(n) > 0$) cannot exceed the number of triples $(x, y, z)$ subject to $x^3 + y^3 + z^3 \leq N$ and $0 \leq x \leq y \leq z$, asymptotically equal to

$$\frac{1}{6}\Gamma(4/3)^3 N = 0.1186\ldots N.$$

Now, a natural question is: does the set of sums of 3 cubes have a density? If so, is it strictly positive? Barrucand [1] computed $f_3(x)$ for $1 \leq x \leq 288000$ and conjectured that it was $o(x)$, as $x$ tends to $\infty$. Vaughan [18] proved in the opposite direction that $f_3(x) \gg_\epsilon x^{8/9-\epsilon}$, improved to $f_3(x) \gg_\epsilon x^{19/21-\epsilon}$ in [19] and then to $f_3(x) \gg_\epsilon x^{11/12-\epsilon}$ in [20] and Hooley [12] conjectured, contrarily to Barrucand, that $f_3(x) \asymp x$. Hooley's approach consists in studying

$$M(x) = \sum_{n \leq x} \mathcal{R}_3(N)^2.$$

He proves a first lower bound

$$M(x) \geq 36 \sum_{n \leq x} \mathcal{R}'_3(n) \sim 6\Gamma(4/3)^3 x,$$

which corresponds to the so-called "combinatorial" contribution, and then a second, taking now into account the "arithmetic" contribution: if

$$F(\theta) = \sum_{n \leq x} \mathcal{R}_3(n) e(n\theta),$$

we have (by just considering the contribution of major arcs)

$$M(x) \geq \int_0^1 |F(\theta)|^2 d\theta \geq \Gamma(4/3)^6 \mathcal{S} x + o(x),$$

with

$$\mathcal{S} = \sum_{q=1}^{\infty} \sum_{\substack{a \bmod q \\ (a,q)=1}} |S(a,q)/q|^6. \tag{2}$$

Hooley conjectures that these two contributions are "independent" and thus that their sum gives the good equivalent for $M(x)$, namely

$$M(x) \sim (6\Gamma(4/3)^3 + \Gamma(4/3)^6 \mathcal{S}) x, \tag{3}$$

which would imply by Cauchy inequality that sums of 3 cubes have a lower density.

## 2. The first probabilistic approach for sums of three cubes

This is due to Erdős and Rényi [8] in 1960. They consider a sequence $(\xi_n)_{n\geq 1}$ of Bernoulli independent random variables such that

$$\Pr(\xi_n = 1) = \alpha_n = \frac{1}{3n^{2/3}}.$$

The random variable counting the number of representations of $N$ as the sum of 3 pseudo-cubes (that is integers $n$ for which $\xi_n = 1$) is

$$R_3(N) = \sum_{\substack{N=h_1+h_2+h_3 \\ h_1<h_2<h_3}} \xi_{h_1}\xi_{h_2}\xi_{h_3}.$$

Erdős and Rényi announced that $R_3(N)$ follows asymptotically a Poisson law:

$$\Pr(R_3(N) = r) \underset{N\to+\infty}{\longrightarrow} \frac{\gamma^r}{r!}e^{-\gamma},$$

where $\gamma = \Gamma(4/3)^3/6$. But their "proof" contained a gap that Landreau [14] recently filled in a general context (cf. [9]) by using original correlation inequalities which also enable him to show that the density of sums of 3 pseudo-cubes is almost surely $1 - e^{-\gamma} = 0.1119\ldots$. This model has the disadvantage to give a positive density for the sums of 2 pseudo-squares, although sums of 2 squares are known to have zero density [13].

## 3. Second probabilistic approach. Sums of three cubes continued

The previous paradox came at least from the following fact: the model did not deal with arithmetic properties of sums of powers. A new model has been recently presented [6] taking into account arithmetic parameters.

Let $K \geq 1$. One builds an integer random sequence $(\mu_l^{(k)})_{l\geq 1}$ restricted to be equal to $k^3$ modulo $K$ and satisfying

$$\mu_l^{(k)} \sim (k + lK)^3$$

almost surely.

Let us denote

$$\rho_3(k, K) = |\{(k_1, k_2, k_3), 1 \leq k_i \leq K : k = k_1^3 + k_2^3 + k_3^3 \bmod K\}|$$

and

$$R_3'(n, K) = |\{n = \mu_{l_1}^{(k_1)} + \mu_{l_2}^{(k_2)} + \mu_{l_3}^{(k_3)}, \mu_{l_1}^{(k_1)} < \mu_{l_2}^{(k_2)} < \mu_{l_3}^{(k_3)}, n = k_1^3 + k_2^3 + k_3^3 \bmod K\}|.$$

Once again, it has been shown that $R_3'(n, K)$ converges in distribution towards a Poisson law:

$$\Pr\{R_3'(n, K) = r\} \underset{\substack{n\to\infty \\ n=k \bmod K}}{\longrightarrow} \frac{1}{r!}\lambda(k, K)^r e^{-\lambda(k,K)},$$

with

$$\lambda(k, K) = \gamma\frac{\rho(k, K)}{K^2}.$$

We can show also that the density of integers such that $R_3'(n, K) > 0$ is almost surely $1 - \delta_0(K)$ where

$$\delta_0(K) = \frac{1}{K}\sum_{k=1}^{K} e^{-\lambda(k,K)}.$$

Notice that it is satisfactory to observe that the probabilistic square mean value satisfies

$$\frac{1}{x}\sum_{n\le x} R'_3(n,K)^2 \sim \Gamma(4/3)^3 + \Gamma(4/3)^6 \mathcal{S}'_2(K)/6$$

which is consistent with Hooley's conjecture (3) (for the definition of $\mathcal{S}'_2(K)$ see the next section).

## 4. Numerical viewpoint.

It is now natural to ask what happens when $K$ tends to infinity. It seems reasonable to consider the following multiplicatively increasing sequence of moduli

$$K_B = \prod_{p^\alpha \le B} p^\alpha.$$

Using convexity of the exponential, we first show that $\delta_0(K_B)$ has a limit $\delta_0$ when $B$ tends to infinity. Then in order to find a good approximation for $\delta_0$, we compute $\delta_0(K_B)$ for a big value of $B$, by developing it in series

$$\delta_0(K_B) = \sum_{i=0}^{I} (-1)^i \frac{\gamma^i}{i!} \mathcal{S}'_i(K_B) + R(K_B),$$

where

$$\mathcal{S}'_i(K_B) = \frac{1}{K_B} \sum_{k \bmod K_B} \left(\frac{\rho(k,K_B)}{K_B^2}\right)^i$$

and

$$|R(K_B)| \le \frac{\gamma^{I+1}}{(I+1)!} \mathcal{S}'_{I+1}(K_B).$$

The multiplicativity of the $\mathcal{S}'_i(K_B)$ is used to estimate them efficiently. Computations have been done using PARI package. For example, with a $B$ around 5000, the truncation parameter $I$ has to be around 18000. Now we use the inequality

$$0 \le \delta_0 - \delta_0(K_B) \le \frac{\gamma^2}{2}(\mathcal{S} - \mathcal{S}'_2(K_B)),$$

where $\mathcal{S}$, defined by equation (2) appears to be also the limit of $\mathcal{S}'_2(K_B)$ as $B$ tends to infinity. Practically, numbers of the form of $K_B$ are replaced by numbers with the following form

$$\prod_{\substack{p^\alpha < B_1 \\ \alpha \ge 2}} p^\alpha \prod_{\substack{p < B_2 \\ p = 1 \bmod 3}} p.$$

This finally allows us to deduce

$$0.09992 \le \delta_0 \le 0.09997.$$

The previous method did not permit to compute $\delta_0$ with an arbitrary number of significant digits. Ph. Flajolet indicated a more efficient method consisting in the use of the Mellin transform. Using the formula

$$e^{-x} = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \Gamma(s) \frac{x^{-s}}{s} ds$$

valid for any $c > 1, x > 0$, we get

$$\delta_0(K) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \Gamma(s) \mathcal{S}'_{-s}(K) \frac{ds}{s} \tag{4}$$

where

$$\mathcal{S}'_{-s}(K) = \frac{1}{K} \sum_{k \bmod K} \left( \frac{\rho(k,K)}{K^2} \right)^{-s}.$$

It then remains to estimate (4) by numerical integration.

## 5. About 7 373 170 279 850

As asserted before, one expects that every sufficiently large integer is $C_4$. Western's conjectures [22] assert that the size of this "last" non-$C_4$ integer has to be in the range between $10^{12}$ and $10^{14}$. Practically, it is intractable to test every integer between $10^{12}$ and $10^{13}$ for example. But the repartition of cubes in arithmetical progressions is far from being regular: this leads to the idea of discriminating the search depending on the class modulo a good integer. So, the strategy has been the following: try to "find" the last non-$C_4$ integer $N_0$ in each class modulo 9. It is considered that it is found if no other non-$C_4$ integer is found between $N_0$ and $10N_0$ (10 is a factor seeming largely sufficient in view of previous experiments). This process allowed to treat the cases of every residue class modulo 9, except 4 and 5. For these ones, it has been needed to proceed to a new discrimination (modulo 7). So there were 14 residue classes modulo 63 to examine. This discrimination has allowed to finish all computations. This has permitted to conjecture that the last non-$C_4$ integer is 7 373 170 279 850; it appears to be equal to 32 modulo 63. Computations have needed 8000 hours. Note that the size of this integer is in accordance with Western's conjectures. This work is more precisely presented in [5].

## Bibliography


[1] Barrucand (Pierre). – Sur la distribution empirique des sommes de trois cubes ou de quatre bicarrés. *Comptes-Rendus de l'Académie des Sciences*, vol. 267, 1968, pp. 409–411.

[2] Bohman (Jan) and Fröberg (Carl-Erik). – Numerical investigation of Waring's problem for cubes. *BIT*, vol. 21, n° 1, 1981, pp. 118–122.

[3] Brüdern (J.). – On Waring's problem for cubes. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 109, n° 2, 1991, pp. 229–256.

[4] Davenport (H.). – On Waring's problem for fourth powers. *Annals of Mathematics*, vol. 40, 1939, pp. 731–747.

[5] Deshouillers (J.-M.), Hennecart (F.), and Landreau (B.). – *7 373 170 279 850.* – Prépublication, UMR 9936, 1996.

[6] Deshouillers (J.-M.), Hennecart (F.), and Landreau (B.). – *Sums of powers: an arithmetic refinement to the probabilistic model of Erdős and Rényi.* – Prépublication, UMR 9936, 1996.

[7] Dickson (L.). – All integers except 23 and 239 are sums of eight cubes. *Bulletin of the American Mathematical Society*, vol. 45, 1939, pp. 588–591.

[8] Erdős (P.) and Rényi (A.). – Additive properties of random sequences of positive integers. *Acta Arithmetica*, vol. 6, 1960, pp. 83–110.

[9] Halberstam (Heini) and Roth (Klaus Friedreich). – *Sequences.* – Springer-Verlag, New York-Berlin, 1983, 2nd edition.

[10] Hardy (G. H.) and Ramanujan (S.). – Asymptotic formulæ in combinatory analysis. *Proceedings of the London Mathematical Society*, vol. 16, 1917, pp. 75–115.

[11] Hilbert (D.). – Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl $n$-ter Potenzen (Waringsche Problem). *Mathematische Annalen*, vol. 67, 1909, pp. 281–300.

[12] Hooley (Christopher). – On some topics connected with Waring's problem. *Journal für die Reine und Angewandte Mathematik*, vol. 369, 1986, pp. 110–153.

[13] Landau (E.). – Über die Einteilung der ... Zahlen in 4 Klassen. *Arch. Math. Phys.*, vol. 13, n° 3, 1908, pp. 305–312.

[14] Landreau (Bernard). – Étude probabiliste des sommes de $s$ puissances $s$-ièmes. *Compositio Mathematica*, vol. 99, n° 1, 1995, pp. 1–31.

[15] Linnik (U. V.). – On the representation of large numbers as sums of seven cubes. *Mat. Sbornik*, vol. 12, n° 54, 1943, pp. 218–224.

[16] McCurley (Kevin S.). – An effective seven cube theorem. *Journal of Number Theory*, vol. 19, n° 2, 1984, pp. 176–183.
[17] Romani (F.). – Computations concerning Waring's problem. *Calcolo*, vol. 19, n° 4, 1982, pp. 415–431.
[18] Vaughan (R. C.). – Sums of three cubes. *Bulletin of the London Mathematical Society*, vol. 17, 1985, pp. 17–20.
[19] Vaughan (R. C.). – On Waring's problem for cubes. *Journal für die Reine und Angewandte Mathematik*, vol. 365, 1986, pp. 122–170.
[20] Vaughan (R. C.). – On Waring's problem for cubes II. *Journal of the London Mathematical Society*, vol. 39, n° 2, 1989, pp. 205–218.
[21] Watson (G. L.). – A proof of the seven cube theorem. *Journal of the London Mathematical Society*, vol. 26, n° 2, 1951, pp. 153–156.
[22] Western (A. E.). – Computations concerning numbers representable by four or five cubes. *Journal of the London Mathematical Society*, vol. 1, n° 2, 1926, pp. 244–251.

# Some Properties of the Cantor Distribution

*Helmut Prodinger*
Technische Universität Wien, Austria

December 9, 1996

[summary by Julien Clément]

**Abstract**

The Cantor distribution is defined as a random series

$$\frac{1-\vartheta}{\vartheta}\sum_{i\ge 1} X_i \vartheta^i,$$

where $\vartheta$ is a parameter and the $X_i$ are random variables that take the values 0 and 1 with probability 1/2. The moments and order statistics are discussed, as well as a "Fibonacci" variation. Connections to certain trees and splitting processes are also mentioned.

## 1. Cantor distribution

**1.1. Random series.** The Cantor distribution with parameter $\vartheta$ ($0 < \vartheta \le 1/2$) was introduced in [5] by the random series

$$X = \frac{\overline{\vartheta}}{\vartheta}\sum_{i\ge 1} X_i \vartheta^i,$$

where the $X_i$ are independent with the distribution $\Pr[X_i = 0] = \Pr[X_i = 1] = \frac{1}{2}$, and $\overline{\vartheta} = 1 - \vartheta$. The name stems from the special case $\vartheta = \frac{1}{3}$, since then this process gives exactly those numbers from the interval $[0, 1]$ that have a ternary expansion solely consisting of the digits 0 and 2. We might alternatively consider an infinite (random) word $w_1 w_2 \cdots$ over the alphabet $\{0, 1\}$ and a map $\mathsf{value}$, defined by

$$\mathsf{value}(w_1 w_2 \cdots) = \frac{\overline{\vartheta}}{\vartheta}\sum_{i\ge 1} w_i \vartheta^i.$$

**1.2. Moments of the distribution.** We abbreviate $a_n = \mathrm{E}[X^n]$. The aim is to solve the recursion formula (from [5])

$$a_n = \frac{1}{2(1-\vartheta^n)}\sum_{k=0}^{n-1}\binom{n}{k}\overline{\vartheta}^{n-k}\vartheta^k a_k, \quad a_0 = 1.$$

Let us introduce the exponential generating function $A(z) = \sum_{n\ge 0} a_n \frac{z^n}{n!}$. The functional equation involving $A(z)$, once solved by iteration, gives

$$A(z) = \prod_{k\ge 0}\frac{1 + e^{\overline{\vartheta}\vartheta^k z}}{2}.$$

In order to derive an asymptotic equivalent of $a_n$, the Poisson generating function $B(z) = e^{-z}A(z)$ has to be considered. Using "Mellin" techniques to derive an asymptotic expansion of $\log B(z)$ when $z$ tends to infinity and a "de-poissonization" argument which suggests the approximation $a_n \sim B(n)$, one gets

$$\mathrm{E}[X^n] = a_n = F(\log_{1/\vartheta} n) n^{-\log_{1/\vartheta} 2} \left(1 + O\left(\frac{1}{n}\right)\right).$$

The function $F(x)$ is periodic of period 1 and has known Fourier coefficients. The mean of $F(x)$ is for instance

$$-\frac{1}{2\log\vartheta} \int_0^\infty \prod_{k\geq 1} \frac{1 + e^{-\overline{\vartheta}\vartheta^k z}}{2} e^{-\overline{\vartheta}x} x^{\log_{1/\vartheta} 2 - 1} dx.$$

**1.3. Order statistics.** Let us consider $n$ random independent variables $Y_1, \ldots, Y_n$ from a Cantor distribution. The average value $\mathrm{E}[\min(Y_1, \ldots, Y_n)]$ of the smallest value among them is denoted by $a_n$. The coefficients $a_n$ obey the following recursion

$$(2^n - 2\vartheta) a_n = \overline{\vartheta} + \vartheta \sum_{k=1}^{n-1} \binom{n}{k} \vartheta a_k.$$

Considering now not exactly the Poisson generating function $A(z) = \sum_{k\geq 0} a_n \frac{z^n}{n!}$ but rather

$$\widehat{A}(z) = \frac{1}{e^z - 1} A(z) = \sum_{n\geq 0} \widehat{a}_n \frac{z^n}{n!},$$

a simpler equation can be obtained. Indeed, one has

$$\widehat{A}(2z) = \vartheta \widehat{A}(z) + \frac{\overline{\vartheta}}{e^z + 1}.$$

The coefficients $\widehat{a}_n$ can be extracted directly from this equation (equating coefficients of $\frac{z^n}{n!}$ on both sides). Going back to the original coefficients $a_n$, we have the explicit solution

$$a_n = -\overline{\vartheta} \sum_{k=0}^{n-1} \binom{n}{k} \frac{B_{k+1}}{k+1} \frac{2^{k+1} - 1}{2^k - \vartheta},$$

where $B_n$ denotes a Bernoulli number. An approach based on Rice's method finally gives an asymptotic equivalent of $a_n$

$$a_n \sim n^{\log_2 \vartheta} \frac{2\vartheta - 1}{\vartheta \log 2} \left(\Gamma(-\log_2 \vartheta)\zeta(-\log_2 \vartheta) + \delta(\log_2 n)\right),$$

where $\zeta(s)$, $\Gamma(s)$ and $\delta(s)$ denote respectively the Riemann's zeta function, the gamma function and a periodic function with period 1 and a very small amplitude (provided $\vartheta$ is not too close to 0).

## 2. Cantor-Fibonacci distribution

**2.1. Fibonacci restriction.** The Cantor distribution might be viewed as a mapping value over a set of random words over a binary alphabet. We might also think about *restricted words*, according to the *Fibonacci restriction*, that two adjacent letters '1' are not allowed. The set of (finite) Fibonacci words $\mathcal{F}$ is given by

$$\mathcal{F} = \{0, 01\}^* \{\epsilon + 1\}.$$

In the original setting (*Cantor distribution*) probabilities are simply introduced by saying that each letter of $\{0, 1\}$ can appear with probability $\frac{1}{2}$. Here the situation is more complicated. We say

that each word of Fibonacci of length $m$ is equally likely. There are $F_{m+2}$ such words, with $F_{m+2}$ denoting the $(m+2)$th Fibonacci number. As an example, consider the classical Cantor case with $\vartheta = \frac{1}{3}$ and $m = 3$. Then the values

$$\mathsf{value}(000) = 0, \quad \mathsf{value}(001) = \frac{2}{27}, \quad \mathsf{value}(010) = \frac{2}{9}, \quad \mathsf{value}(100) = \frac{2}{3}, \quad \mathsf{value}(101) = \frac{20}{27}$$

appear, each with probability $\frac{1}{5}$. The generating function $F(z)$ of Fibonacci words, according to their lengths is easily derived from the definition of $\mathcal{F}$ above,

$$F(z) = \frac{1+z}{1-z-z^2} = \sum_{m\geq 0} F_{m+2} z^m.$$

Note that

$$F_n = \frac{1}{\sqrt{5}}(\alpha^n - \beta^n) \qquad \text{with} \qquad \alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2}.$$

**2.2. Moments of the Cantor-Fibonacci distribution.** Let us consider the generating functions

$$G_n(z) := \sum_{w\in\mathcal{F}} (\mathsf{value}(w))^n \, z^{|w|},$$

where $|w|$ denotes the length of the Fibonacci word $w$. The quantity

$$\frac{[z^m]G_n(z)}{[z^m]F(z)}$$

is the $n$th moment, when considering words of length $m$. Then we let $m$ tend to infinity to get a limit called $M_n$ (note that taking limits wasn't necessary for the independent original case). The recursion for $\mathsf{value}$, when restricted to Fibonacci words, is

$$\mathsf{value}(0w) = \vartheta \cdot \mathsf{value}(w)$$
$$\mathsf{value}(10w) = \overline{\vartheta} + \vartheta^2 \cdot \mathsf{value}(w).$$

These formulae translate almost directly to generating functions according to the recursive definition $\mathcal{F} = \epsilon + 1 + \{0, 10\}\mathcal{F}$. Thus it gives an explicit recursion formula for the functions $G_n(z)$

$$G_n(z) = \frac{1}{1-\overline{\vartheta}^n z - \vartheta^{2n} z^2}\left[\overline{\vartheta}^n z + z^2 \sum_{i=0}^{n-1} \binom{n}{i} \overline{\vartheta}^{n-i} \vartheta^{2i} G_i(z)\right].$$

Since we only consider the limit for $m \to \infty$, we can get the asymptotic behaviour noting that both $F(z)$ and $G_n(z)$ have the same dominant singularity at $z = 1/\alpha$ and also that it is a simple pole. Consequently, we have (due to a "pole cancellation")

$$M_n = \lim_{m\to\infty} \frac{[z^m]G^n(z)}{[z^m]F(z)} = \lim_{z\to 1/\alpha} \frac{G_n(z)}{F(z)}.$$

Therefore we have the following theorem

THEOREM 1. *The moments of the Cantor-Fibonacci distribution fulfill the following recursion:* $M_0 = 0$ *and for* $n \geq 1$

$$M_n = \frac{1}{\alpha^2 - \alpha\vartheta^n - \vartheta^{2n}} \sum_{i=1}^{n} \binom{n}{i} \overline{\vartheta}^{n-i} \vartheta^{2i} M_i.$$

**2.3. The asymptotic behaviour of the moments.** A rough estimate shows that $M_n \approx \lambda^n$. We might infer that $\lambda = \overline{\vartheta} + \lambda\vartheta^2$, so that $\lambda = \frac{1}{1+\vartheta}$. It is not rigourous but we can set

$$m_n := M_n \cdot (1+\vartheta)^n$$

anyway and show that this sequence has nicer properties. As before the recurrence on the coefficients $m_n$ and then the exponential generating function $m(z) = \sum_n m_n \frac{z^n}{n!}$ need to be considered. Finally the Poisson transformed function $\widehat{m}(z) = e^{-z}m(z)$ obeys the functional equation

$$\widehat{m}(z) = \frac{e^{-\overline{\vartheta}z}}{\alpha}\widehat{m}(\vartheta z) + \frac{1}{\alpha^2}\widehat{m}(\vartheta^2 z).$$

Because $m_n \sim \widehat{m}(n)$, the next step considers the behaviour of $\widehat{m}(z)$ for $z \to \infty$. Using the Mellin transform (and the Mellin inversion formula), we have the following theorem

THEOREM 2. *The nth moment $M_n$ of the Cantor-Fibonacci distribution has for $n \to \infty$ the following asymptotic behaviour*

$$M_n = \left(1+\overline{\vartheta}\right)^{-n} \Phi(-\log_\vartheta n) n^{\log_\vartheta \alpha} \left(1 + O\left(\frac{1}{n}\right)\right),$$

*where $\Phi(x)$ is a periodic function with period 1 and known Fourier coefficients. The mean (zeroth Fourier coefficient) is given by*

$$-\frac{1}{\log\vartheta}\int_0^\infty \frac{e^{-\overline{\vartheta}z}}{\alpha}\widehat{m}(\vartheta z) z^{-\log_\vartheta \alpha - 1} dz.$$

Note that here, $\frac{e^{-\overline{\vartheta}z}}{\alpha}\widehat{m}(\vartheta z)$ is merely considered as an auxiliary function. This integral can be computed numerically by replacing $\widehat{m}(\vartheta z)$ by the first few values of its Taylor expansion, which can be obtained through the recursion formula on the coefficient $m_n$. As an example, the classical case $\vartheta = \frac{1}{3}$ gives (apart from small fluctuations),

$$M_n \sim .6160498 n^{-.4380178} 0.75^n.$$

The fact that in an asymptotic formula the generating function itself, evaluated at a certain point, appears is not at all uncommon in combinatorial analysis.

## Bibliography


[1] Flajolet (Philippe), Gourdon (Xavier), and Dumas (Philippe). – Mellin transforms and asymptotics: harmonic sums. *Theoretical Computer Science*, vol. 144, n° 1-2, 1995, pp. 3–58. – Special volume on mathematical analysis of algorithms.

[2] Flajolet (Philippe) and Sedgewick (Robert). – Mellin transforms and asymptotics: finite differences and Rice's integrals. *Theoretical Computer Science*, vol. 144, n° 1-2, 1995, pp. 101–124. – Special volume on mathematical analysis of algorithms.

[3] Grabner (P. J.) and Prodinger (H.). – Asymptotic analysis of the moments of the Cantor distribution. *Statistics & Probability Letters*, vol. 26, n° 3, 1996, pp. 243–248.

[4] Knopfmacher (Arnold) and Prodinger (Helmut). – Explicit and asymptotic formulae for the expected values of the order statistics of the Cantor distribution. *Statistics & Probability Letters*, vol. 27, n° 2, 1996, pp. 189–194.

[5] Lad (F. R.) and Taylor (W. F. C.). – The moments of the Cantor distribution. *Statistics & Probability Letters*, vol. 13, n° 4, 1992, pp. 307–310.

[6] Prodinger (H.). – The Cantor-Fibonacci distribution. *Applications of Fibonacci numbers*, 1998. – To appear.

# Graph Colouring via the Probabilistic Method

*Bruce Reed*

Équipe Combinatoire CNRS
Université Pierre et Marie Curie, Paris, France

April 21, 1997

[summary by François Morain and Philippe Robert]

## 1. Introduction

Colouring a graph with the minimum number of colours is a classical problem in graph theory and has many applications. For instance, think of a cellular phone network on which each vertex (phone) must use a different frequency with its neighbours. This problem is also known to be a difficult one (see for instance [3]).

The purpose of the talk is to present a naive algorithm for colouring a certain type of graphs and explain how to analyze it with elementary probabilistic tools that we will describe first.

## 2. Probabilistic tools

Throughout this section we denote by $\Pr(A)$ the probability of the event $A$, $E(X)$ the expected value of the random variable $X$, and $E(X/A_1, \ldots, A_n)$ the conditional expectation of $X$ relative to the events $A_1, \ldots, A_n$.

**2.1. The Lovász Local lemma.** Suppose that on some probability space $\Omega$, there are $n$ events $A_1, \ldots, A_n$ that are undesirable. We wish to estimate if there is a positive probability to avoid any of them, i.e., if there is a positive lower bound for the quantity

$$\Delta = \Pr\left(\cap_{i=1}^{n} A_i^c\right),$$

where $A_i^c = \Omega - A_i$. If the events are independent, that is for any $k$-tuple $1 \leq i_1 < \cdots < i_k \leq n$,

$$\Pr(\cap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} \Pr(A_{i_j}),$$

then

$$\Delta = \prod_{1=1}^{n} (1 - \Pr(A_i)).$$

The problem is that in practice, the events are not always completely independent but weakly independent, in the sense that for each $i$ there exists a subset $V_i \subset \{1, \ldots, n\}$ such that $A_i$ is independent of the events $A_j, j \in V_j^c$. In other words, $A_i$ is possibly dependent of $A_j$ with $j$ in the "neighbourhood" $V_i$ of $i$. If the cardinality of the $V_i$'s is small, one might expect an estimate close to the one we saw for the independent case. This is the conclusion of Lovács's lemma, see [1].

LEMMA 1. *If the events are such that for all* $1 \leq i \leq n$,

1. $\Pr(A_i) \leq p$,
2. *$A_i$ is independent of $(A_j)_{j \notin V_i}$*,
3. $|V_i| \leq d$,

*and if $ep(d+1) < 1$ then none of the events $A_i$, $i = 1, \ldots, n$, occurs with positive probability.*

**2.2. Azuma's inequality.** If $(Y_i)$ be a sequence of independent random variables with the same distribution on $\{0,1\}$, $p = \Pr(Y_i = 1)$; The result of successive coin tossings is a good model for this sequence of random variables. It is well known that the time averages $\frac{1}{n}\sum_{i=1}^{n} Y_i$ converges exponentially fast to $p$ as $n \to +\infty$. Rigourously, this is Chernoff's bound

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - p\right| > a\right) < 2e^{-a^2/3np},$$

it says basically that with high probability $[p/a^2]$ coin tossings are sufficient to get an estimate of $p$ with an accuracy of the order of $a$. This kind of result has been extended, for independent variables, to the case of arbitrary distributions, i.e., not only with values in $\{0,1\}$, as long as they have an exponential moment. This is a part of large deviations theory, see [2].

Another possible generalization is to consider the case where instead of the sum of independent variables, one looks at some functional $X$ of some arbitrary random variables $Y_1, \ldots, Y_n$ with values in $\{0,1\}$. Azuma's inequality says that if the conditional exceptions of $X$ with respect to $Y_1, \ldots, Y_i$ do not jump sharply as $i$ goes from 1 to $n$, then $X$ is concentrated around its average value, formally,

PROPOSITION 1. *If for each $i \leq n$,*

$$\max_{y_1,\ldots,y_{i+1} \in \{0,1\}} |E(X/Y_1 = y_1, \ldots, Y_i = y_i, Y_{i+1} = y_{i+1}) - E(X/Y_1 = y_1, \ldots, Y_i = y_i)| \leq c_i, \quad (1)$$

*then*

$$\Pr(|X - E(X)| > a) \leq 2e^{-\frac{a^2}{2\sum_1^n c_i^2}}.$$

Azuma's inequality is surprisingly sharp considering the weak hypotheses of the proposition. In the independent case, for $X = \sum_1^n Y_i$, condition (1) is satisfied with $c_i = 1$, hence the inequality is in this case,

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - p\right| > a\right) < 2e^{-\frac{-a^2}{2n}},$$

which is very close to Chernoff's bound.

## 3. Graph colouring

We *colour* a graph $G$ such that every pair of adjacent vertices receive different colours. The *chromatic number* of $G$, noted $\chi(G)$ is the minimum number of colours required to colour $G$. It is easy to see that, if $\Delta(G)$ denotes the maximal degree of $G$, then $\chi(G) \leq \Delta(G) + 1$.

We can obtain good bounds for $\chi(G)$ for certain types of graphs, as explained in [4]. For fixed $\varepsilon > 0$, we saw that a vertex $v$ is *$\varepsilon$-sparse* if the subgraph induced by $N_v$, the neighbourhood of $v$, has at most $(1-\varepsilon)\frac{\Delta(\Delta-1)}{2}$ edges. A graph is *$\varepsilon$-sparse* if each of its vertices is $\varepsilon$-sparse.

THEOREM 1. *For $\Delta$ sufficiently large, if $G$ has maximum degree $\Delta$ and $G$ is $\varepsilon$-sparse, then $\chi(G) \leq (1 - \varepsilon/2e^6)\Delta$.*

Let us indicate a rough proof of this theorem. In a first step, we construct a partial colouring $\mathcal{C}$ of $G$ such for each vertex $v$, the number of neighbours of $v$ which are coloured exceeds the number of colours appearing on $N_v$ by at least $\frac{\varepsilon}{2e^6}\Delta + 1$.

From this, we complete the colouring of $\mathcal{C}$ to a $\left(1 - \frac{\varepsilon}{2e^6}\right)\Delta$-colouring of $G$ in a greedy manner: We colour the remaining vertices one at a time. When we come to colour $v$, there must be an available colour: Since $v$ has at most $\Delta$ neighbours (this is where the sparseness comes in), the number of colours appearing in $N_v$ is bounded by

$$\Delta - \left(\frac{\varepsilon}{2e^6}\Delta + 1\right).$$

Hence fewer than $\left(1 - \frac{\varepsilon}{2e^6}\right)\Delta$ colours appear in its neighbourhood.

Let us come back to the construction of $\mathcal{C}$. We first assign each vertex of $G$ a uniformly random colour from $\{1, 2, \ldots, \lceil \Delta/2 \rceil\}$. If two adjacent vertices have the same colour, we uncolour them. The resulting partial colouring yields $\mathcal{C}$.

The first thing to show is that $\mathcal{C}$ is not too small, which is rather easy. Then we must study, for vertex $v$, the random variable $Z_v$ which counts the number of pairs of vertices in $N_v$ which have the same colour in $\mathcal{C}$. It can be shown that since $G$ is sparse, the expectation of $Z_v$ is greater than $\varepsilon\Delta/e^4$.

Now that we have proved that many vertices in $N_v$ are coloured, we must show that $Z_v$ does not differ too much from its expected value. Once this is done, we use the Local Lemma to prove that *every* vertex will have such a property, thus proving the property on $\mathcal{C}$. By a technical argument replacing $Z_v$ with a more amenable quantity, Azuma's Inequality is used to prove the assumption on $Z_v$. Roughly speaking, the idea is that a colouring of $v$ should not influence the colouring of the other parts of $\mathcal{C}$, since $G$ is sparse.

## Bibliography


[1] Alon (Noga) and Spencer (Joel H.). – *The probabilistic method.* – John Wiley & Sons Inc., New York, 1992, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, xvi+254p.

[2] Dembo (Amir) and Zeitouni (Ofer). – *Large deviations techniques and applications.* – Jones and Bartlett Publishers, Boston, MA, 1993, xiv+346p.

[3] Karp (Richard M.). – Reducibility among combinatorial problems. In *Complexity of Computer Computations.* pp. 85–103. – New York, 1972. Proceedings of a Symposium held at IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., 1972.

[4] Molloy (M.) and Reed (B.). – Graph colouring via the probabilistic method. – April 1997. Preprint.

# Part 5

# Miscellany

# The Dynamics of Continued Fractions with Periodic Constraints

*Brigitte Vallée*
University of Caen

June 9, 1997

[summary by Philippe Flajolet]

**Abstract**

Consider rational, quadratic, or real numbers whose continued fraction representations satisfy periodic constraints. A typical instance is numbers whose continued fraction quotients are alternatively odd and even. Such sets have a somewhat fractal nature, and the Hausdorff dimension of the set of reals as well as the density of the set of rationals satisfying such constraints can be determined. Other consequences include a precise analysis of the height of constrained continued fractions. The methods rely on a transfer operator that generates the constraints and whose dominant spectral properties prove essential.

## 1. Introduction

The triadic Cantor set $\mathcal{C}$ formed with numbers whose ternary representation does not contain the digit 2 is perhaps the most ancient instance of a set defined by constrained number representations. The density of triadic rationals $a/3^n$ that belong to $\mathcal{C}$ is clearly $(2/3)^n$ and this set has a fractal Hausdorff dimension equal to $\log_3 2$; see for instance [1].

Continued fractions are of course a well-studied representation system and it is tempting to consider similarly what happens when constraints are imposed on them. The problems are more delicate since digits in continued fraction expansions are not independent in the common probabilistic sense. *Elementary* constraints are defined by a single set $M \subset \mathbb{N}$ and one imposes that all continued fraction digits (quotients) should belong to $M$. I. J. Good initiated this line of study in 1941 for $M = \{1, 2\}$, while recent results have been obtained by Hensley [3] for finite sets $M$.

Here, more general types of constraints are studied. Let $\ell \geq 1$ be an integer called the *period length*, and $\mathcal{M} = M_1 \times \cdots \times M_\ell$ a family of sets (noted multiplicatively) called the *period*. A number $x \in [0, 1]$ will be said to be $\mathcal{M}$-constrained if the sequence of its continued fraction digits if of type $M_1, M_2, \ldots, M_\ell, M_1, M_2, \ldots, M_\ell, \ldots$, cyclically. For instance, we have

$$e - 2 = \exp(1) - 2 = /1, 2, 1, 1, 4, 1, 1, 6, 1, \ldots/ = \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{\cdots}}}},$$

so that $e - 2$ obeys (for instance) the constraint $\mathcal{M} = \{1\} \times \{2, 4, 6, \ldots\} \times \{1\}$ of length 3. Such sets arise naturally in connection with the robustness of hashing functions in cryptography [7].

## 2. Transfer Operators

The works of Mayer, Hensley, and Vallée (see, e.g., [2] for a gentle introduction) have demonstrated the importance of *transfer operators* in this range of problems. Let

$$U(x) = \frac{1}{x} - [\frac{1}{x}]$$

be the so-called continued-fraction *shift*. Classically, many features of the basic continued fraction algorithm and Euclid's algorithm are captured by an operator (taken to act on families of analytic functions),

$$\mathcal{G}_s[f](z) := \sum_{m\geq 1} (\frac{1}{m+z})^s f(\frac{1}{m+z}).$$

The operator $\mathcal{G}_s$ is such that its specialization $\mathcal{G}_2[f]$ represents the probability density that results from applying one stage of the shift $U$ to a random number drawn with initial density $f$. (This results from a simple argument that traces all the possible antecedents of a given real number.) It is known, since the works of Lévy, Kuzmin, and Wirsing, that properties of this operator, for instance its spectrum, are closely related to metric properties of continued fractions; see for instance [5, 6].

In the context of constrained numbers, it is then natural to introduce operators associated to an elementary constraint,

$$\mathcal{G}_{M,s}[f](z) := \sum_{m\in M} (\frac{1}{m+z})^s f(\frac{1}{m+z}),$$

and, next in order of complexity, to a family of periodic constraints

$$\mathcal{G}_{\mathcal{M},s} := \mathcal{G}_{M_\ell,s} \circ \mathcal{G}_{M_{\ell-1},\ s} \circ \cdots \circ \mathcal{G}_{M_1,s}.$$

These operators play a rôle analogous to generating functions in analytic combinatorics. Algebraically (or formally), they generate all constrained continued fractions. By design, $\mathcal{G}_{M,s}$ generates all linear fractional transformations (LFT's) of depth 1 that are constrained by $M$, the $r$th iterate $\mathcal{G}^r_{M,s}$ similarly generates all the LFT's of depth $k$, and, for instance, the quasi-inverse

$$\omega(s) = (I - \mathcal{G}_{\mathcal{M},s})^{-1}[1](0)$$

yields the Dirichlet series of all rational numbers (of arbitrary depth) constrained by $\mathcal{M}$.

Analytically, density properties of constrained rationals are related to values of $s$ such that $I - \mathcal{G}_{\mathcal{M},s}$ ceases to be invertible, since then objects like $\omega(s)$ become singular. Clearly, such singular values are determined by values of $s$ such that 1 is an eigenvalue of $\mathcal{G}_{\mathcal{M},s}$.

The operator $\mathcal{G}_{\mathcal{M},s}$ is known to be compact (and even "nuclear" in the terminology of Grothendieck), which implies that its spectrum is discrete and the eigenvalues are "separated" from each other. In addition, for $s$ real, the operator satisfies a strong positivity property visible from its definition. Thus, a generalized Perron-Frobenius theory—of which Markov chains and positive matrices are a very particular instance—applies, and there is a unique *dominant* eigenvalue that is simple and positive, $\lambda(\mathcal{M}, s)$. Then, for a wide class of constraints (technically, of the so-called "open type"), the domain of existence of the function $s \mapsto \lambda(\mathcal{M}, s)$ is an open set, and there exists a unique real $s_{\mathcal{M}}$ for which

$$\lambda(\mathcal{M}, s_{\mathcal{M}}) = 1.$$

This number $s_{\mathcal{M}}$ is then a dominant singularity of the function $s \to (I - \mathcal{G}_{\mathcal{M},s})^{-1}[1](0)$ and its rôle is essential.

In a way, the situation here is reminiscent of the analysis of the combinatorial *sequence schema*, $h(z) = (1 - g(z))^{-1}$ in the "supercritical case", where $g(z)$ attains the value 1 before becoming singular.

## 3. Results

The number $s_{\mathcal{M}}$ always belongs to the interval $[0,2]$. A first batch of results deals with metric information on constrained numbers, expressed in terms of the special quantity $s_{\mathcal{M}}$.

($i$) The Hausdorff dimension of the set $\mathbb{R}(\mathcal{M})$ of the $\mathcal{M}$-constrained reals is equal to $\frac{1}{2}s_{\mathcal{M}}$;

($ii$) The density of the set $\mathbb{Q}(\mathcal{M})$ of the $\mathcal{M}$-constrained rationals is equal to $s_{\mathcal{M}}$. This means that the subset of constrained rationals $p/q$ satisfying $1 \le p < q \le N$ and $\gcd(p,q) = 1$ has a cardinality that satisfies

$$|\mathcal{Q}_N(\mathcal{M})| \sim c_{\mathcal{M}}\, N^{s_{\mathcal{M}}};$$

($iii$) A similar result holds for constrained quadratic irrationals of size $\le N$. There, size is measured naturally by the smallest fundamental solution of Pell's equation, $x^2 - \Delta y^2 = 4$, with $\Delta$ the associated discriminant.

These three results have counterparts regarding characteristics of continued fraction expansions of constrained numbers, and especially the length of expansions, a parameter of great interest in the analysis of algorithms like Euclid's.

($iv$) The mean length of the continued fraction representation of a rational in $Q_N(\mathcal{M})$ (i.e., a rational constrained by $\mathcal{M}$ whose denominator is $\le N$) satisfies

$$\mathrm{E}[X_N(\mathcal{M})] \sim \frac{1}{\mathcal{L}_{\mathcal{M}}} \log N,$$

where $\mathcal{L}_{\mathcal{M}}$ is the so-called Lévy constant associated with the constraints $\mathcal{M}$;

($v$) Similar results hold for constrained irrationals of "size" $\le N$.

The value of Lévy's constant for unconstrained irrationals is

$$\mathcal{L} = \frac{\pi^2}{12 \log 2},$$

and, accordingly, this constant occurs in the expected number of steps of Euclid's algorithm, asymptotically $\frac{1}{\mathcal{L}} \log N$, a well-known result of Heilbronn and Dixon. In the general context of constraints, its value is related to the dominant eigenvalue $\lambda(\mathcal{M}, s)$ and to the "critical value" $s_{\mathcal{M}}$ by

$$\mathcal{L}_{\mathcal{M}} = -\frac{1}{\ell}\, \frac{d}{ds} \lambda(\mathcal{M}, s)\Big|_{s=s_{\mathcal{M}}}.$$

Thus, the statement ($iv$) vastly generalizes the Heilbronn-Dixon analysis.

The proofs rely on an algebraic description of sets and parameters of interest by means of the $\mathcal{G}_{\mathcal{M},s}$ operators, as already explained. Then spectral properties (using positivity and Perron-Frobenius properties, as well as compactness) play an essential rôle. Finally, quantitative estimates are derived by means of Tauberian theorems applied to relevant Dirichlet series. The methods thus constitute an interesting parallel to those of analytic combinatorics, the problems being naturally more delicate, as one has to appeal to functional analysis and coefficient extraction of Dirichlet series. It is worth mentioning also that the constants appearing here can be estimated to reasonable accuracy (10 digits say), by means of truncation methods that have proved instrumental in the estimation of Wirsing's constant or Vallée's constant.

Interesting questions are also suggested by this talk that is to be presented at the *Journées Arithmétiques*, Limoges, September 1997. Along distributional lines, Hensley [4] proved in 1994 that the number of steps of Euclid's algorithm is asymptotically Gaussian. Probably, similar properties hold for constrained numbers. Can this be proved (e.g., by operator methods and simple perturbation theory)? Also, what about other types of constraints like "*no two quotients equal to* 1 *in a row*"? For strings, we know that the density is about $(\phi/2)^n$ with $\phi$ the golden section. It

is tempting to conjecture the existence of a fractal dimension for reals and a related density for rationals, of the form $N^{\alpha}$. More generally, is there a theory of "regular constraints" that would be the counterpart for continued fractions of regular languages in the realm of strings?

## Bibliography


[1] Falconer (K. J.). – *The Geometry of Fractal Sets.* – Cambridge University Press, 1986, *Cambridge Tracts in Mathematics*, vol. 85, xiv+162p.

[2] Flajolet (Philippe) and Vallée (Brigitte). – *Continued Fraction Algorithms, Functional Operators, and Structure Constants.* – Research Report n° 2931, Institut National de Recherche en Informatique et en Automatique, July 1996. 33 pages. Invited lecture at the 7th Fibonacci Conference, Graz, July 1996; to appear in *Theoretical Computer Science.*

[3] Hensley (Doug). – Continued fraction Cantor sets, Hausdorff dimension, and functional analysis. *Journal of Number Theory*, vol. 40, n° 3, 1992, pp. 336–358.

[4] Hensley (Doug). – The number of steps in the Euclidean algorithm. *Journal of Number Theory*, vol. 49, n° 2, 1994, pp. 142–182.

[5] Knuth (Donald E.). – *The Art of Computer Programming.* – Addison-Wesley, 1981, 2nd edition, vol. 2: Seminumerical Algorithms.

[6] Rockett (Andrew M.) and Szüsz (Peter). – *Continued Fractions.* – World Scientific Publishing Co. Inc., River Edge, NJ, 1992, x+188p.

[7] Tillich (Jean-Pierre) and Zémor (Gilles). – Hashing with $SL_2$. In *Advances in Cryptology. Lecture Notes in Computer Science*, vol. 839, pp. 40–49. – Berlin, 1994. Proceedings CRYPTO '94, Santa Barbara, CA, 1994.

# A History of Cryptology

*François Morain*

LIX, École polytechnique

April 21, 1997

[summary by Pierrick Gaudry]

**Abstract**

Cryptology is very old but has got a renewal of interest. Until the end of the 80's, it was reserved to military people or diplomats, but it is now accessible to the general public. Cryptology contains the art of hiding information and the techniques to break a secret. This talk is a brief survey of the history of this art.

## Introduction

First of all, one needs to recall the precise meaning of some terms. *Cryptography* is the art of communicating confidentially through an insecure channel. *Cryptanalysis* is the art of deciphering those communications when one is not the legitimate receiver. And *Cryptology* is the union of these two domains.

Two basic principles are known for cryptography: *substitution*, which consists in permuting the letters of the alphabet, and *transposition*, which permutes the letters of the text.

## 1. From prehistory to the modern era

During antiquity, writing was safe because only a few people could read. We can however note some simple substitutions in India and the use of special or rare symbols by scribes in Mesopotamia. In Greece and Rome, the use of cryptography increased for military purposes. In Sparta, in 475 B.C, was invented the *scytale*, which is a conic stick around which one encircles a strip of paper, and then writes the message vertically. Julius Caesar used a simple substitution system.

After this period and till the fifteenth century, the only valuable cryptographic activity was in the Arabic civilization. Qalqahandi wrote an encyclopedia with a section dedicated to cryptology, with the first appearance of cryptanalysis.

Cryptology came back in occident with the Renaissance (first in Italy). A lot of techniques appear at this time, notably polyalphabetical substitutions. The principle is to have a key (for example CADEAU), and to "add" the repeated key to the message we want to encrypt.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | c | r | y | p | t | o | g | r | a | p | h | e |
| + | c | a | d | e | a | u | c | a | d | e | a | u |
| = | E | R | A | T | T | I | I | R | D | T | H | W |

In France, during the Renaissance, the monarchy used a system of nomenclator; there was a quite good security thanks to the frequent change of code and to the existence of spare codes.

## 2. From telegraph to radio

In the middle of the nineteenth century, the telegraph generated a new craze for cryptography and cryptanalysis. Kasiski in 1863 gave a method to attack polyalphabetical substitutions. Mathematics are introduced in cryptology, and Kerckhoffs gave a few laws that should be verified by a "good" cryptographic system. In particular he insists on the fact that the system has to be public, the only secret being a key.

During World War I, England decrypted a lot of German messages. A decisive one was the Zimmermann telegram which proved the double game played by Germany with Mexico and the USA. The publication of this telegram in the American press incited the USA to go to war.

In France, a remarkable cryptanalysis was achieved by Painvin. He broke the German system ADFGX in April 1918. Before they launched their last offensive, the Germans modified the system, and in a few days Painvin broke it once more. The French then discovered where Ludendorff wanted to attack, and could stop the offensive.

## 3. The automation of cryptology

The most important invention of the beginning of the twentieth century is the *one-time-pad* by Vernam. The principle is to "add" an infinite random sequence to the message. The problem is then to build a pseudo-random generator. That was done with the invention of the rotor which gives a polyalphabetical substitution with a huge period. All the machines used during World War II were based on this principle.

Japan first used the RED machine which was broken by classical spying, and then the PURPLE system which was cryptanalysed by Friedman in 1940. Germany used the famous ENIGMA machine, regularly improved by addition of cabling and a choice of three rotors between five. There was two great centers of cryptanalysis: the first in USA with Friedman, and the second in England with Turing and Welchman. In the latter, a lot of German messages where decrypted. The principle of the cryptanalysis was to take advantage of some weaknesses of the usage that Germans did of ENIGMA: they had a very strict format for the beginning of the messages, and some operators did not choose random keys.

## 4. The last fifty years

The key facts of the last years are the increasing development of computers, and the great interest of civilians for cryptography.

In the beginning of the 70's, the National Bureau of Standards decided to publish a cryptosystem which could be used by governmental agencies or banks; this was done in 1977, with the Data Encryption Standard (DES). Concurrently to this, Diffie invented in 1975 the concept of public-key cryptosystem, which was applied by Rivest, Shamir and Adleman (RSA) in 1977.

There is now a great link between cryptology and some branches of modern mathematics and computer science: probability theory, information theory, algorithmic number theory, or the theory of error correcting codes are useful tools.

Nowadays, some new applications of cryptography appear: electronic trade, money, or notarial deeds.

## Bibliography

[1] Kahn (David). – *The codebreakers; the story of secret writing.* – Macmillan, New York, 1967.

[2] Schneier (Bruce). – *Applied cryptography: protocols, algorithms, and source code in C.* – Wiley, New York, 1996, 2nd edition.

# Contents

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS
Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
(France)
http://www.inria.fr
ISSN 0249-6399